CUH-H06410-9-26146

The British Journal for the Philosophy of Science

BPS

C.

Cambridge University Press

Feb

May 1966-67

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

Volume XVII Part 1 15s. net

The British Journal for the Philosophy of Science

VOLUME XVII Part 1 MAY 1966

CONTENTS

ARTICLES

THEORY AND OBSERVATION (I) 1
by Marshall Spector

A LOGIC FOR EVIDENTIAL SUPPORT (I) 21
by L. Jonathan Cohen

COULD A MACHINE PERCEIVE? 44
by Alan Gauld

NOTES AND COMMENTS

A PARADOX OF INFORMATION 59
by David Miller

A COMMENT ON MILLER'S NEW PARADOX OF INFORMATION 61
by Karl R. Popper

FUNCTIONAL ANALYSES IN BIOLOGY 69
by Harry G. Frankfurt and Brian Poole

CONTENTS—(*contd.*)

REVIEWS

The Discovery of Time. By Stephen Toulmin and June Goodfield reviewed by R. M. Young 73

Beyond the Edge of Certainty. Edited by Robert G. Colodny reviewed by L. Laudan 76

The Age of Automation. By Sir Leon Bagrit reviewed by R. Needham 80

Concepts and Principles in the Space-Time Theory within Einstein's Special Theory of Relativity. By H. Törnebohm reviewed by G. F. R. Ellis 83

Methodology of the Behavioral Sciences. By Rollo Handy reviewed by A. W. Still 85

Plato, The Midwife's Apprentice. By I. M. Crombie reviewed by G. E. R. Lloyd 87

OBITUARY 88

© Cambridge University Press, 1966

BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

The British Society for the Philosophy of Science is a body of which the purpose is described by the following statement from its Constitution and Rules:

> 'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. The annual subscription is 10s. 6d. Those who wish to join should write to the Hon. Secretary, Mr J. W. N. Watkins, London School of Economics, Houghton Street, London, W.C.2.

The Northern Group of the British Society for the Philosophy of Science meets regularly at various centres in the north of England. Those who wish to join it should write to Dr W. Mays, Hon. Secretary of the Northern Group of the British Society for the Philosophy of Science, Dept. of Philosophy, The University, Manchester, 13.

The Present Committee of the Society is as follows:

Professor S. Körner
(President)

Dr G. J. Whitrow
(President-Elect)

Mr J. W. N. Watkins
(Hon. Secretary)

Dr C. W. Kilmister
(Hon. Treasurer)

Mr G. Buchdahl

Dr H. D. Sluga

Dr J. S. Wilkie

Dr W. Mays
(representing Northern Group)

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

The *British Journal for the Philosophy of Science* is the official organ of the Society. It is published quarterly (February, May, August and November) by Cambridge University Press; it may be obtained from any bookseller or direct from the publisher. The annual subscription is 42s. (U.S.A. $7.50), post free; single issues 15s. (U.S.A. $2.75) plus postage. The subscription is 31s. 6d. for members of the British Society for the Philosophy of Science who should order direct from the publishers.

Articles and correspondence for inclusion in the *Journal* should be sent to the Editor, Dr Mary Hesse, Whipple Museum, Free School Lane, Cambridge. Contributions should normally not exceed 5,000 words in length. With an article a summary of about 100 words (which should also be used as an abstract) should be sent. **A sheet giving instruction about lay-out, style of references, etc., can be obtained from the Editor on request.**

Exchange of Abstracts. When a paper is accepted for publication, the abstract will then be published in *Dialectica*, *Methodos* and *Philosophy of Science*. Likewise, abstracts of papers in these journals will be found in our pages.

Publishers are invited to forward copies of books to the Editor.

Particulars regarding advertising in the *Journal* may be obtained from Cambridge University Press, Bentley House, 200 Euston Road, London, N.W.1; or in USA and Canada, Cambridge University Press American Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A.

The British Journal for the Philosophy of Science

Volume 17 No. 2 AUGUST 1966

CONTENTS

ARTICLES

Theory and Observation (II) 89
by Marshall Spector

A Logic for Evidential Support (II) 105
by L. Jonathan Cohen

The Use of Statistics in Explanation 127
by Arthur W. Collins

NOTE

A Paradox of Zero Information 141
by Karl R. Popper

DISCUSSIONS

Miller's So-called Paradox of Information 144
y J. L. Mackie

On a So-called So-called Paradox 147
y David Miller

rigin and Concept of Relativity (Parts I and II) 149
y G. H. Keswani

REVIEWS

he Retreat to Commitment. By William W. Bartley III 153
viewed by Peter Jones

ilosophy of Science: The Delaware Seminar. Vol. 2. Edited by Bernard Baumrin 155
iewed by David Hirschmann

rces and Fields. By Mary B. Hesse 159
iewed by L. Pearce Williams

m and Strategy in Science. Edited by John R. Grigg and F. T. C. Harris 160
iewed by William Kneale

CONTENTS—(*contd.*)

The Explanation of Behaviour. By Charles Taylor reviewed by Margaret A. Boden 162

Recognition: A Study in the Philosophy of Artificial Intelligence. By Kenneth M. Sayre reviewed by Max Clowes 165

RECENT PUBLICATIONS 168

ABSTRACTS 174

ANNOUNCEMENT ii

ANNOUNCEMENT

The Netherlands Society for Logic and Philosophy of Science announces that the Third International Congress for Logic, Methodology and Philosophy of Science will be held in Amsterdam from 25 August to 2 September 1967 under the auspices of the Division of Logic, Metholodogy and Philosophy of Science of the International Union of History and Philosophy of Science. The Proceedings will be organised into the following sections:

1. Mathematical Logic
2. Foundations of Mathematical Theories
3. Automata and Programming Languages
4. Philosophy of Logic and Mathematics
5. General problems of Methodology and Philosophy of Science
6. Methodology and Philosophy of Physical Sciences
7. Methodology and Philosophy of Biological Sciences
8. Methodology and Philosophy of Psychological Sciences
9. Methodology and Philosophy of Social Sciences
10. Methodology and Philosophy of Linguistics
11. History of Logic, Methodology and Philosophy of Science

Symposia will be organised in several sections.

Further information for those willing to attend or to submit papers available from:

The Secretariat, 3rd International Congress for Logic, Methodolog
and Philosophy of Science, c/o Holland Organizing Centre, 16 Lang
Voorhout, The Hague, Netherlands.

© Cambridge University Press, 1966

BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

HE British Society for the Philosophy of Science is a body of which the urpose is described by the following statement from its Constitution and Rules:

> 'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once month during the academic year. These are held, by kind permission, in niversity College, London.

Ordinary membership is open to all persons who have been proposed y two members, approved by the Committee, and elected at an ordinary meeting. The annual subscription is 10s. 6d. Those who wish to join ould write to the Hon. Secretary, Mr J. W. N. Watkins, London School Economics, Houghton Street, London, W.C.2.

The Northern Group of the British Society for the Philosophy of Science eets regularly at various centres in the north of England. Those who wish join it should write to Dr W. Mays, Hon. Secretary of the Northern oup of the British Society for the Philosophy of Science, Dept. of Philophy, The University, Manchester, 13.

The Present Committee of the Society is as follows:

Professor S. Körner
(President)

Dr G. J. Whitrow
(President-Elect)

Mr J. W. N. Watkins
(Hon. Secretary)

Dr C. W. Kilmister
(Hon. Treasurer)

Mr G. Buchdahl

Dr H. D. Sluga

Dr J. S. Wilkie

Dr W. Mays
(representing Northern Group)

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

The *British Journal for the Philosophy of Science* is the official organ of th Society. It is published quarterly (February, May, August and Novembe by Cambridge University Press; it may be obtained from any bookseller c direct from the publisher. The annual subscription is 42s. (U.S.A. $7.5 post free; single issues 15s. (U.S.A. $2.75) plus postage. The subscription 31s. 6d. for members of the British Society for the Philosophy of Science wh should order direct from the publishers.

Articles and correspondence for inclusion in the *Journal* should be se to the Editor, Dr Mary Hesse, Whipple Museum, Free School Lan Cambridge. Contributions should normally not exceed 5,000 words i length. With an article a summary of about 100 words (which should als be used as an abstract) should be sent. **A sheet giving instruction abou lay-out, style of references, etc., can be obtained from the Editor o request.**

Exchange of Abstracts. When a paper is accepted for publication, t abstract will then be published in *Dialectica, Methodos* and *Philosophy Science.* Likewise, abstracts of papers in these journals will be found in pages.

Publishers are invited to forward copies of books to the Editor.

Particulars regarding advertising in the *Journal* may be obtained fro Cambridge University Press, Bentley House, 200 Euston Road, Londo N.W.1; or in USA and Canada, Cambridge University Press Americ Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A.

The British Journal for the Philosophy of Science

VOLUME 17 No. 3 NOVEMBER 1966

CONTENTS

ARTICLES

THE *RAISON D'ÊTRE* OF INDUCTIVE ARGUMENT 177
by Max Black

A NEW LOOK AT THE PROBLEM OF INNATE IDEAS 205
by Nicholas Rescher

THE ACHILLES PARADOX AND TRANSFINITE NUMBERS 219
by C. David Gruender

DISCUSSIONS

CAN A MACHINE BE CONSCIOUS? 232
by Robert J. Clack

ORIGIN AND CONCEPT OF RELATIVITY 234
by G. H. Keswani

REVIEWS

The Structure of Science. By Ernest Nagel 237
Article-review by Paul K. Feyerabend

The Basic Laws of Arithmetic. By Gottlieb Frege 249
Translated and edited, with an Introduction
by Montgomery Firth
A Study of Frege. By Jeremy D. B. Walker
reviewed by Aaron Sloman

Philosophical Problems of Space and Time. By Adolf Grünbaum 253
reviewed by David Bohm

'Towards an Historiography of Science.' By Joseph Agassi 256
reviewed by Thomas S. Kuhn

CONTENTS—(*contd.*)

Hobbes's System of Ideas. By J. W. N. Watkins reviewed by Samuel I. Mintz 259

The Foundations of Metaphysics in Science. By Errol E. Harris reviewed by W. H. Thorpe 261

Studies in Plato's Metaphysics. Edited by R. E. Allen reviewed by G. E. R. Lloyd 263

ERRATUM 264

ANNOUNCEMENT 264

BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

HE British Society for the Philosophy of Science is a body of which the rpose is described by the following statement from its Constitution and ules:

> 'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'.

The Society holds meetings, at which papers are read and discussed, once month during the academic year. These are held, by kind permission, in niversity College, London.

Ordinary membership is open to all persons who have been proposed y two members, approved by the Committee, and elected at an ordinary eeting. The annual subscription is 10s. 6d. Those who wish to join ould write to the Hon. Secretary, Professor J. W. N. Watkins, London chool of Economics, Houghton Street, London, W.C.2.

The Northern Group of the British Society for the Philosophy of Science eets regularly at various centres in the north of England. Those who wish o join it should write to Dr W. Mays, Hon. Secretary of the Northern roup of the British Society for the Philosophy of Science, Dept. of Philophy, The University, Manchester, 13.

The present Committee of the Society is as follows:

Professor S. Körner
(President)

Dr G. J. Whitrow
(President-Elect)

Professor J. W. N. Watkins
(Hon. Secretary)

Dr C. W. Kilmister
(Hon. Treasurer)

Mr G. Buchdahl

Dr H. D. Sluga

Dr J. S. Wilkie

Dr W. Mays
(representing Northern Group)

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

The *British Journal for the Philosophy of Science* is the official organ of th Society. It is published quarterly (February, May, August and November by Cambridge University Press; it may be obtained from any bookseller o direct from the publisher. The annual subscription is 42s. (U.S.A. $7.50) post free; single issues 15s. (U.S.A. $2.75) plus postage. The subscription i 31s. 6d. for members of the British Society for the Philosophy of Science wh should order direct from the publishers.

Articles and discussions for inclusion in the *Journal* should be sen to the Editor, Dr Mary Hesse, Whipple Museum, Free School Lan Cambridge. Contributions should normally not exceed 5,000 words i length. With an article a summary of about 100 words (which should als be used as an abstract) should be sent. **A sheet giving instruction abou lay-out, style of references, etc., can be obtained from the Editor o request.**

Exchange of Abstracts. When a paper is accepted for publication, ou abstract will then be published in *Dialectica, Methodos* and *Philosophy of Science.* Likewise, abstracts of papers in these journals will be found in the pages of this *Journal.*

Publishers are invited to forward copies of books to the Editor.

Particulars regarding advertising in the *Journal* may be obtained fr Cambridge University Press, Bentley House, 200 Euston Road, Lond N.W.1; or in USA and Canada, Cambridge University Press Americ Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A.

The British Journal for the Philosophy of Science

VOLUME 17 NO. 4 FEBRUARY 1967

CONTENTS

ARTICLES

ome Relations Between Philosophical and Scientific Theories 265
S. Körner

efutation or Comparison? 279
G. C. Archibald

n Models 297
J.W. Swanson

aresis and the Alleged Asymmetry Between Explanation and Prediction 313
y Paul Dietl

NOTES AND COMMENTS

On the Principle of Total Evidence 319
by I. J. Good

e White Shoe is a Red Herring 322
I. J. Good

REVIEWS

Basic Concepts of Measurement. By Brian Ellis 323
eviewed by D. H. Mellor

Philosophic Foundations of Quantum Mechanics. By H. Reichenbach 326
eviewed by Paul K. Feyerabend

Analytical Philosophy of History. By Arthur C. Danto 328
eviewed by Patrick Gardiner

olitical Argument. By B. Barry 331
viewed by R. S. Peters

CONTENTS—(*contd.*)

The Stratification of Behaviour. By D. S. Shwayder 33
reviewed by D. W. Hamlyn

The Pursuit of Certainty. By Shirley Robin Letwin 3
reviewed by John Dunn

Philosophische Bemerkungen. By Ludwig Wittgenstein 3
reviewed by H. D. Sluga

The Voices of Time. Edited by J. T. Fraser 3
reviewed by J. D. North

Science and Philosophy. By Vincent Edward Smith 3
reviewed by William J. Shea

Logic, Methodology and Philosophy of Science. Edited by Yehoshua Bar-Hillel 3
reviewed by R. Harré

Introduction to Logic and to the Methodology of the Deductive Sciences. By Alfred Tarski 34
reviewed by James Cargile

ANNOUNCEMENTS 34

ABSTRACTS 348

CORRIGENDUM 3

RITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

itish Society for the Philosophy of Science is a body of which the is described by the following statement from its Constitution and

The purpose of the Society is to study the logic, the method and the osophy of science as well as those of the various special sciences, including ocial sciences.'

Society holds meetings, at which papers are read and discussed, once during the academic year. These are held, by kind permission, in ity College, London.

nary membership is open to all persons who have been proposed members, approved by the Committee, and elected at an ordinary g. The annual subscription is 10s. 6d. Those who wish to join write to the Hon. Secretary, Professor J. W. N. Watkins, London of Economics, Houghton Street, London, W.C.2.

Northern Group of the British Society for the Philosophy of Science egularly at various centres in the north of England. Those who wish it should write to Dr W. Mays, Hon. Secretary of the Northern f the British Society for the Philosophy of Science, Dept. of Philoe University, Manchester, 13.

esent Committee of the Society is as follows:

ssor S. Körner
(President)

. J. Whitrow
(President-Elect)

fessor J. W. N. Watkins
(Hon. Secretary)

fessor C. W. Kilmister
(Hon. Treasurer)

Mr G. Buchdahl

Dr H. D. Sluga

Dr J. S. Wilkie

Dr W. Mays
(representing Northern Group)

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

The *British Journal for the Philosophy of Science* is the official orga Society. It is published quarterly (February, May, August and No by Cambridge University Press; it may be obtained from any book direct from the publisher. The annual subscription is 42s. (U.S.A. post free; single issues 15s. (U.S.A. $2.75) plus postage. The subsc 31s. 6d. for members of the British Society for the Philosophy of Scie should order direct from the publishers.

Articles and discussions for inclusion in the *Journal* should to the Editor, Dr Mary Hesse, Whipple Museum, Free Scho Cambridge. Contributions should normally not exceed 5,000 w length. With an article a summary of about 100 words (which sh be used as an abstract) should be sent. **A sheet giving instructio lay-out, style of references, etc., can be obtained from the Ed request.**

Exchange of Abstracts. When a paper is accepted for publicati abstract will then be published in *Dialectica*, *Methodos* and *Philos Science.* Likewise, abstracts of papers in these journals will be found pages of this *Journal.*

Publishers are invited to forward copies of books to the Editor.

Particulars regarding advertising in the *Journal* may be obta Cambridge University Press, Bentley House, 200 Euston Road N.W.1; or in USA and Canada, Cambridge University Press Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A.

The British Journal for the
Philosophy of Science

VOLUME XVII MAY, 1966 PART 1

THEORY AND OBSERVATION (I)*

MARSHALL SPECTOR

Introduction

this paper I wish to examine certain basic assumptions of an impor-
t currently held view as to the nature of physical theories. Accord-
to this view, physical theories are to be analysed or reconstructed as
piriically interpreted logistic systems. A theory is said to have two
mponents or aspects: its calculus, which is the logical skeleton of the
ory, looked upon as devoid of empirical meaning, and the empirical
erpretation of the calculus, the set of 'direct meanings' or 'semanti-
rules' which are given for (some of) the non-logical constants
curring in the calculus. Not all of the non-logical constants need be
ectly interpreted via semantical rules; those which are not are said to
tain an indirect and partial meaning if they occur in sentences of the
ory which also contain terms which are directly interpreted.
ese sentences are referred to as 'correspondence rules'.) These
ns are also said to gain their empirical meaning in the entire context
e theory in which they occur.

thin this general framework of analysis several assumptions are
First, there is what I shall call the 'dual language assumption'.
stated by Carnap as follows:

> ept] the customary and useful [division of] the language of
> to two parts, the observation language and the theoretical
> [, where] the observation language uses terms designating
> properties for the description of observable things and events
> theoretical language [uses] terms which may refer to unob-
> nts, unobservable aspects or features of events, e.g. to micro-
> 1

.vi.65

nap, 'The Methodological Character of Theoretical Concepts', in
s *in the Philosophy of Science*, I (Minneapolis, University of Minnesota

That is, it is assumed that in looking at an unreconstructed theory—a theory as it may be found in a textbook or treatise, we can distinguish in general two classes of terms, the observation terms and the theoretical (non-observation) terms. ‘ Theoretical ’ and ‘ observation(al) ’ are taken as denoting mutually exclusive categories of terms (as well as the non-linguistic entities to which they refer—so-called observable and theoretical objects, properties, events, etc.)

Secondly, it is assumed that when we come to reconstruct a theor in the manner advocated above, those non-logical constants of t calculus which represent the observation terms of the unreconstruct theory are the only ones which can be given semantical rules. I maintained that *because* the theoretical terms purportedly refe entities which cannot be (directly) observed, such terms cannot directly understood and so cannot be given semantical rules in completed reconstruction of the theory. Such terms can be underst *only* in the context of the entire theory in which they occur.

This latter assumption characterises what I shall call the ‘ par interpretation ’ version of the interpreted logistic system analysis; th retical terms are given only an indirect and partial interpretat through the rôle they play in the theory as a whole, rather than direc by semantical rules.[1]

In the paper mentioned in the footnote below, I attempted to ref the second assumption and cast doubt on the first, by means of an direct argument based on an analysis of the concept of a model fo physical theory. In the present paper, I shall attempt to refute b of these assumptions in a more direct manner—by an examinatio the concepts ‘ observation term ’ and ‘ theoretical term ’ (and the re concepts ‘ observable entity ’, and ‘ theoretical entity ’).

Part A will be devoted to an examination of the conce observation term. First, I shall consider some of the attempts this concept which have been offered by proponents of interpretation thesis ’. Two conclusions will become none of the analyses offered will capture the pre-analy these authors seem to have in mind as evidenced by the give of observation terms and observable entities; (ii) analyses will be such that it will follow from them that terms *cannot* be given semantical rules when reconstruc in which they are found. Secondly, I shall show that the terms 'observable' and 'unobservable', in some of their important common uses, will not provide the type of distinction the partial interpretation theorists require either. Points (i) and (ii) will be seen to apply to these common uses as well as to the actual analyses of the concepts in question given by the partial interpretation theorists. It will become apparent that the terms 'observable' and 'unobservable' are commonly used in several distinct ways, to make different points about a term or entity. No one of these will be such that all and only those terms which the authors in question have *called* observable will turn out to be observable, nor will they be such that direct meanings can be given only for those terms which turn out to be observational.

---

[1] For a more complete description of this type of analysis, see t my paper ‘ Models and Theories ’, *British Journal for the Philosophy o* 1965, 121–142.

In Part B, I will offer the same considerations for the concept of a theoretical term. (Thus it should be apparent that I will also conclude that 'theoretical' and 'observable' do not denote mutually exclusive categories.)

A. The Concept of an Observation Term

1. *The Partial Interpretation Theorists' Analysis*

Let us begin by considering what Carnap has written about the concept of observability. In his more recent works, he has not given an analysis of the notion of an observation term, but instead has drawn an admittedly rough distinction between observation terms and non-observation terms on the basis of the types of designata which terms may have. Observation terms are those which designate observable things and their observable properties, and these turn out to be, by stipulation, only gross physical objects and their observable properties (with the latter applying only to the former). As Carnap puts it, material objects which are '. . . neither too small nor too large for an observer to see them and take them in his hands'.[1] The observation language of science is, *by definition*, a portion of ordinary language—what Carnap has called the 'thing language'. (Thus, if we were to ask whether unobservable objects could be characterised by observational predicates—a question which we have found to be of some importance in my paper mentioned above and have answered in the affirmative—the answer would *have to be No*, and not because this is the way our theories have in fact developed, but because of the very nature of the concept of an unobservable object. Similarly, it would follow that unobservable properties could not characterise observable objects, which is false on its face.)

---

[1] Loc. cit. p. 38. If this is taken literally, the moon would have to be considered unobservable, as would the Earth.

Braithwaite does no better in making clear what he means by an observation term. The notion of ' observation ', he says, is to be taken in its ' wide ' or ' ordinary ' sense. The *analysis* of this notion, he tells us, is an epistemological problem which need not concern philosophers of science. ' Observed facts are usually facts about the behaviour of material bodies.'[1]

In some of his earlier works, Carnap maintained that the notion of an observation term is a *pragmatic* one (as opposed to syntactical or semantical). That is, whether a term in a language is an observation term depends on whether users of that language employ the term in a certain manner—as reactions to and descriptions of their immediate experience. In *Foundations of Logic and Mathematics* (p. 203), he wrote:

> We find among the concepts of physics—and likewise among those of the whole of empirical science—differences of abstractness. Some are more elementary than others, in the sense that we can apply them in concrete cases on the basis of observations in a more direct way than others. The others are more abstract; in order to find out whether they hold in a certain case, we have to carry out a more complex procedure, which, however, finally rests on observations.

He also claimed that there is no sharp line between ' abstract ' terms and ' elementary ' terms, but his analysis proceeded as if there were. (Elementary terms *have* semantical rules which are understood; abstract terms simply do not.[2]) In ' Testability and Meaning ', he gave the following pragmatic characterisation of the notion of an observation term (for predicates):

> A predicate ' P ' of a language L is called *observable* for an organism (e.g. a person) N, if, for suitable arguments, e.g. ' b ', N is able to *come to a decision with the help of a few observations about a full sentence, say ' P(b) '*

---

[1] Braithwaite, *Scientific Explanation*, p. 4.

[2] Hempel (' The Theoretician's Dilemma ', *Minnesota Studies in the Philosophy of Science*, 3, 42-43) and Nagel (*The Structure of Science*, p. 83) also say that the distinction may not be a sharp one—that there may be borderline cases. But the partial interpretation version of the interpreted logistic system analysis does not provide a means for reconstructing this fact, if it is a fact. For observation terms are those which can be given semantical rules—direct interpretations, while theoretical terms *cannot*. The best that could be done would be to *draw* a line by convention. But one cannot do justice to the fact that a sharp lines does not exist (assuming this to be the case) by saying that we can draw one wherever we wish, and that moreover we *must* draw one, and then treat the categories distinguished in quite different ways.

> [my emphasis], i.e. to a confirmation of either ' P(b) ' or ' -P(b) ' of such a high degree that he will either accept or reject ' P(b) '.... There is no sharp line between observable and non-observable predicates because a person will be more or less able to *decide a certain sentence quickly* [my emphasis].... For the sake of simplicity we ... draw a sharp distinction.[1]

(Notice that the clause ' for suitable arguments ' does allow the possibility of unobservable *objects* having (some) observational *properties*, for the requirement is only that there be *some* arguments for which the predicate meets the condition—not that the application of the predicate to all possible arguments meets the condition.)

Hempel also has given a pragmatic characterisation of the observational vocabulary of science (though a different one than that given by Carnap). He writes:

> The condition ... imposed upon the observational vocabulary is of a pragmatic character; it demands that each term included in that vocabulary be of such a kind that under suitable conditions, different observers can, by means of direct observation, *arrive at a high degree of agreement on whether* the term applies [my emphasis] to a given situation. The expression ' coincides with ' ... is an example.[2]

More recently, Grover Maxwell has offered a pragmatic characterisation of the concept of an observation term. Maxwell spends some time in trying to make his characterisation clear and precise, and draws some general conclusions from his analysis regarding the ontological status of unobservable objects. In light of this, and since Braithwaite, Carnap, and Hempel have little more to say about the notion of observability than that which I have quoted above, it will be worthwhile to take a close look at Maxwell's proposed explication.

Maxwell writes as follows:

> We should take [as the unit of the observation base of science] not the ' observation term ' but, rather, the quickly decidable sentence.... A quickly decidable sentence (in the technical sense employed here) may be defined as a singular, non-analytic sentence such that a reliable, reasonable sophisticated language user can very quickly decide whether to assert it or deny it when he is reporting on an occurrent situation.

---

[1] Reprinted in Feigl and Brodbeck (eds.), *Readings in the Philosophy of Science*, p. 63.

[2] Hempel, ' A Logical Appraisal of Operationism ', in Philipp Frank (ed.), *The Validation of Scientific Theories*, New York, Collier Books, 1961, p. 67.

> 'Observation term' may now be defined as a 'descriptive (nonlogical) term which may occur in a quickly decidable sentence'. . . .[1]

As a footnote to 'very quickly decide', he adds:

> We may say 'noninferentially decide', provided this is interpreted liberally enough to avoid starting the entire controversy about observability all over again. [ibid.]

It is not clear that deciding 'very quickly' and deciding 'noninferentially' are the same. Some inferences may be drawn quite quickly, and some decisions, although not made quickly, may involve no process of inference. But I shall not press this point; in what follows, I shall assume that either would be acceptable to Maxwell.

This characterisation does not rule out the possibility that one may use inference (based on a theory, perhaps) to *justify* the acceptance of a sentence of the kind described. This can always be asked for, although there are situations where such a request would be deemed silly, unless the person asking for the justification had something out of the ordinary in mind. (E.g. 'How do you know that the sentence "There is a table before me", which you just quickly accepted, is true? I understand that a jocular psychologist has just rigged this room.') However, I think it must also be required that a person who accepts a sentence in the manner described by Maxwell must be prepared to draw conclusions from the sentence so asserted in order to justify or test the claim made in light of the theory (or conceptual scheme) in which it is embedded. Furthermore, these 'tests' must bear out the truth of the statement. After all, I can quickly decide to accept any sentence whatsoever. As Maxwell's characterisation stands, without this qualification, it actually purports to define 'observation-*like*' terms. (Compare the usual distinction between 'law-like' sentences and 'laws'.) This amounts to a requirement of *reliability*, and is similar to Carnap's requirement of a high degree of confirmation in the passage quoted from him a few pages back. I do not think this characterisation is adequate for the purposes of the partial interpretation theorists; but before I argue this, let us examine what appear to be some of its merits.

---

[1] Grover Maxwell, 'The Ontological Status of Theoretical Entities', in *Minnesota Studies in the Philosophy of Science*, 3, 13. See also his 'Theories, Frameworks, and Ontology', *Philosophy of Science*, April 1962; and Paul Feyerabend's 'An Attempt at a Realistic Interpretation of Experience', *Proceedings of the Aristotelian Society*, 1957, 58, 143.

I do not mean to imply that Maxwell advocates the partial interpretation version of the interpreted logistic system analysis of physical theories. I shall be considering his analysis of observability in detail because it is similar to and more carefully presented than those used by Carnap and Hempel. (Feyerabend quite definitely rejects the partial interpretation thesis.)

The characterisation does not, by itself, say anything about the specific *content* of the observational vocabulary of science. It does not pick out terms to be considered as observation terms on the basis of what sorts of *designata* they may have. The history of logical empiricism could well be told in terms of the problems and changes involved in characterising the ' observation language ' of science on the basis of the *content* of the observational vocabulary—what the terms refer to: e.g. subjective impressions, inter-subjective sense data, and finally, gross physical objects and their properties. Maxwell's characterisation does not decide this issue; it makes no assumptions as to *what sorts* of entities can be observed. Instead, it purports to provide an analysis of the concept of an observation term. In so doing, it allows as observable entities those objects and properties which Braithwaite would say are ' ordinarily ' classified as such. Thus, terms like ' table ' meet the condition of Maxwell's characterisation, as do ' brown ', ' hard ', and even terms like ' ornate '. (After all, is it not correct to say that we simply *see* that this is an ornate table?) The observation language fulfilling Maxwell's characterisation would be (in fact, and not by stipulation) the descriptive parts of ordinary language—mainly the ' thing language ', containing terms referring to material objects, their ' ordinary ' properties, and their relations to one another. *Most* of the terms which *turn out to be* observation terms on the basis of the pragmatic characterisation are those which are stipulated to be observation terms by Carnap and Braithwaite.[1]

But it is quite important to notice that it is not *only* the thing language which fulfils this pragmatic characterisation. For example, ' lightning bolt ', ' spark ', ' flame ', ' rainbow ', and ' shadow ' are observation terms in the sense of Maxwell's characterisation, even though these terms do not refer to ' physical objects ' or their properties. This is as it should be; an analysis of the concept of an observation term which rules these terms out of the observational vocabulary would be unacceptable. The term ' wind ' would also be an observation term, notwithstanding the possible objection that we never ' really observe the wind itself' since it is in fact transparent and colourless. We *can* quickly and noninferentially decide in certain situations whether a sentence such as ' the wind is blowing through the trees ' is to be accepted—by looking or listening. Winds just are not the sorts of entities that have a colour (just as a rainbow does not have a texture).

---

[1] It also seems to capture the sense in which science is said to rest on observations, both as a stimulus for theorising and as a court of appeal for theories once constructed (Braithwaite's ' observable facts '). This might be called the functional sense of ' observation term '; it sketches generally the *rôle* of an observation language when *used as* part of a theory (or, the observational rôle of a term when used in relation to a theory).

I should like to examine the wind example more closely in order to sharpen the pragmatic characterisation of observability—in particular, to see whether for the purpose of reconstructing theories as interpreted logistic systems, we should take it as a concept relative to *each individual observer*, to some *standard observer* (as Maxwell does), or relative to what is *possible in principle.* It will develop that Maxwell's characterisation cannot be accepted, that is, it will not provide the type of distinction required by the partial interpretation theorists. The aspects of it which seemed worthy of merit will also lead to its unacceptability. I shall begin by considering the concept relative to individual observers.

Imagine a person who has been raised indoors where care has been taken never to expose him to air currents (because of some rare skin disease, for example). When he is first taken outdoors he may have to be told that the (for him) strange sensations he feels over the surface of his body are (or, are caused by) the wind blowing against him. (Similarly, for the noises he hears in the trees, etc.) He may, for a time, be forced to *infer* from these sensations to the presence of something else—the wind—on the basis of the physics he has learned during his confinement. For this person, ' wind ' would not be an observation term. It would not occur in any singular, non-analytic sentence which *he* could quickly decide whether to accept or reject. But eventually, it will become one, in the sense of Maxwell's characterisation (and in the senses given by Carnap and Hempel in the passages quoted from them). He will eventually become ' good at ' being able to tell quickly that it is windy (say); he will eventually be able quickly (and correctly) to decide whether to apply or withhold terms like ' stiff breeze '. A term which was once *not* an observation term for him and will have *become* one. He will have learned to observe the wind.[1]

But if this is correct, it would seem that the line between what is observed and what is not (and thus must be inferred) depends on how much the observer in question knows (or on the amount of ' training ' he has), and will also vary in time for a given observer. Thus, we could not draw a *general* distinction between two types of terms—observational and non-observational. Maxwell, however, does not relativise the notion of an observation term to the individual observer. Instead, he draws the observable-unobservable distinction for a *standard observer*, which he describes as the ' reliable, reasonably sophisticated language user '. (Thus the term ' wind ', which would not be an observation term for the person of the example, would still be classified as an observation term for purposes of reconstructing physical theories.) If we relativise the notion to the standard observer in this way, it is still possible that a term may be non-observational at one time (for Maxwell's standard observer) and later become observational. This has in fact happened in the history of science, as Maxwell himself has pointed out.

---

[1] An often cited example of this sort of situation is the case of the newly sighted having to learn how to quickly and noninferentially apply what are for us observation terms. Such a person has to learn to be able to see even physical objects.

Perhaps the most common example of this is the term 'molecule ', since large molecules can now be observed (in Maxwell's sense) with the aid of the electron microscope. Another example is the term ' germ ' (or ' virus ') where the concept of a germ was first introduced into science as referring to unobservable entities (' disease entities '). It might be objected that the standard observer, even if identified with Maxwell's ' reasonably sophisticated language user ', cannot quickly decide (or cannot decide noninferentially) that ' what he sees ' by use of a microscope or other instrument is in fact the same thing he has read about in the newspaper—' virus X ', say. He has to be taught to ' interpret the shadows he really sees '[1] in light of a quick course in microbiology and microscopy, and might very well not recognise what is before him when presented with different though similar specimens. *But he can be taught to observe these things* (i.e. he can be taught to quickly decide . . .), just as in the case of the wind example, the person had to be taught to observe the wind.

This fact introduces us to the first of two important defects which this characterisation of observability as relativised to the *standard observer* has. In the first place, it is difficult to see at just what level of actual ignorance (or lack of ' sophistication ') the line is to be drawn between the standard observer and the observer who has had ' special training '. As Paul Feyerabend has so aptly said (in a different context):

---

[1] See Gustav Bergmann, 'An Empiricist Philosophy of Physics ', reprinted in Feigl and Brodbeck, *Readings in the Philosophy of Science*, p. 287: ' . . . when I look through a microscope or a telescope, all I see is a patch of colour which creeps through the field like a shadow over a wall.'

> By the way, who *is* this uninstructed layman? For the purpose for which he is being employed in many arguments, it would seem to emerge that he is not supposed to know much about science, or much politics, or much religion, or much of anything. This means that in these times of mass communication and mass education he must be very careful not to leave his television set on for too long a time, and he must also not allow himself to converse too much with his friends, his children, etc. That is, he must either be a savage or an idiot.[1]

Perhaps Maxwell's reasonably sophisticated language user is supposed to know more than the ' uninstructed layman ' that Feyerabend makes light of, but the following point still holds: The fact that there is a comparatively small number of people (scientists—and this would vary from science to science) for whom certain terms are observational *in Maxwell's sense* is merely a contingent fact of no theoretical significance. Anyone *can learn* enough biology, say, for the term ' virus ' to become observational for him—to be able to decide quickly that there is a virus before him on the stage of the electron microscope. The relativisation of the concept of observability to some standard observer is unstable. There is no sharp line between the standard observer and the observer with special training; and even if there were, it would shift with time as science, technology, and the general level of education advanced. The term ' wind ' is not considered to be a non-observation term for purposes of theory reconstruction just because we can think of situations where someone would have to acquire some ' special training ' before it became observational for him. Why then should we do this for the term ' virus ' (say) when the same sort of thing obtains? A general analysis of the structure of physical theory which bases its fundamental distinctions on the observation that most people have only a certain amount of the knowledge that *is available* to them, and have only this small amount as a matter of empirical fact rather than in principle, is to be rejected or severely modified.

The second defect which I have in mind is more serious. Suppose we were to grant that Maxwell's characterisation relativised to the standard observer *is* adequate to distinguish those terms which Maxwell and the other authors in question *call* non-observational from those which are *called* observational. The resulting distinction, I maintain, would not be strong enough to support the different treatment given to the two categories so distinguished, by the partial interpretation reconstruction of physical theories. For the difference between observation and non-observation terms, as these notions are characterised by Maxwell, is of a *pragmatic* character. It depends on whether or not the standard observer is able (as a matter of empirical fact) to do certain things with a term—decide quickly whether a certain type of sentence containing it is to be accepted. But the partial interpretation thesis rests upon the claim that the difference between observation terms and non-observation terms is of a *semantical* nature; it assumes that observation terms and non-observation terms are so essentially *different in meaning* that we *cannot* give semantical rules for non-observation terms in the reconstruction of a theory containing such terms. This is at best an *assumption* that would need justification, but the partial interpretation theorists attempt no such justification. My claim is that the assumption *cannot* be justified merely on the basis of Maxwell's characterisation of an observation term. The distinction between quickly decidable terms and non-quickly decidable terms does not *necessitate* giving semantical rules only to the former.

---

[1] Paul Feyerabend, 'Explanation, Reduction, and Empiricism ', *Minnesota Studies in the Philosophy of Science*, 3, 83 n.

Suppose, then, that we were to drop the relativisation to the *standard observer*, and accept Maxwell's characterisation relative to what the *physicist* (biologist, etc.) is able to do. That is, a term would be called an observation term if it occurred in a singular, non-analytic sentence for which it is possible for a person who knows all that *is* known about the designata of the term (e.g. knows the physical theories which contain the term, and their applications) to decide quickly (or non-inferentially) whether to accept or reject the sentence. (This would seem to be the most natural modification that could be made to take into account the objections I have just raised.) If Maxwell's characterisation is modified in this way, there will be (almost) *no distinction* that could be drawn in accordance with it, for *almost all terms* in all well-confirmed physical theories will have to be classified as observational. A physicist can decide very quickly, and without employing any inference, whether or not an electron has passed through a cloud chamber before him. He can also decide very quickly (in the affirmative) whether there is a *gravitational field* pervading his laboratory. And if he notices a certain series of pulses on the screen of an oscilloscope which is connected in a certain manner with a scintillation counter placed near a nuclear reactor, he will very quickly decide that the sentence 'A neutrino has just passed through the scintillation counter' is to be accepted.[1] And yet the authors in question would want to call electrons and neutrinos *unobservable.* (Maxwell grants that gravitational fields may be observable.) Their writings would indicate, in fact, that these are almost paradigm cases of unobservable entities.

Finally, there is nothing to prevent the interested layman (especially if he is a 'sophisticated language user') from *learning* enough of the theories involved to be able to so the same things. We may conclude, therefore, that Maxwell's characterisation cannot be accepted in any of the various forms in which we have considered it. (Carnap's and Hempel's characterisations clearly meet the same fate.) The only sense in which it can support any distinction at all is one which is of no theoretical significance, and which cannot support the partial interpretation analysis which is built upon it, in which it is claimed that non-observation terms cannot be directly understood.

In the remainder of Part A of this paper, I should like to examine some of the points that could be, and actually have been made by referring to an object as observable or unobservable. I shall attempt to establish two points: (1) There is no *single* distinction that can be drawn which will reflect some *general* distinction between types of objects—observable and unobservable; (2) no *one* of the distinctions that *could* be drawn will capture the distinction that the partial interpretation theorists *intend.* (Here I have in mind what I called (i) and (ii) toward the end of the introduction.)

The several uses of the 'observable-unobservable' terminology that I shall be considering are those which actually occur in ordinary situations—including ordinary scientific situations; and yet the authors in question have not attempted to distinguish them. This is a fair criticism, I believe, because quite often their comments on observability would lead one to believe that they are using the term 'observable' and 'unobservable' in their ordinary sense. Braithwaite, in fact, explicitly claims that this is what he is doing. He writes:

> In this book . . . the experience to which scientific laws are referred will be taken as including all observable facts, where observation is taken in its wide, everyday meaning and with no philosophical limitation.[2]

---

[1] Cf. Philip Morrison, 'The Neutrino', *Scientific American*, January 1956. (With respect to the characterisations of an observation term given by Carnap and Hempel, which were quoted earlier in the paper, it should be noticed that his fellow physicists would quickly *agree* with Morrison, and that the sentence would be quite *highly confirmed.*)

[2] *Scientific Explanation*, p. 4.

2. *Factors Involved in the Use of Instruments*

In this section, I shall restrict myself to various considerations that arise from the use of instruments. Both Carnap and Maxwell have suggested that a distinction can be drawn between observable and unobservable (or only ' indirectly ' observable) objects on the basis of whether or not instruments are (or must be) used in making an observation.[1] Let us see whether this can provide a way of drawing the distinction intended. Carnap writes:

> . . . the predicate ' electric field of such and such an amount ' is not observable to anybody, because, although we know how to test a full sentence of this predicate, we cannot do it *directly, i.e. by a few observations; we have to apply certain instruments* and hence to make a *great many preliminary observations* in order to find out whether the things before us are instruments of the kind required. [My emphasis.][2]

Now I think that there are *several* interesting distinctions that *can* and *should* be drawn on the basis of considerations involving the use of instruments; but the requirement that there be only a few ' preliminary observations ' for an object to be classified as observable cannot plausibly be said to be one of them. For a physicist does not have to ' make a great many preliminary observations in order to find out whether ' the object before him is his laboratory galvanometer (say); thus, the electric current in a wire that he measures with his instrument would have to be classified as observable. Yet Carnap would presumably maintain that electric currents are unobservable. Nor will the relative *complexity* of instruments used to make observations serve to draw the distinctions intended by Carnap. An electroscope is a very simple instrument, and with its aid we can determine (with very few preliminary observations) whether there is an electric charge on a body; yet

---

[1] If the pragmatic characterisation offered by Maxwell were to be accepted, these further considerations would not be necessary.

The reader may detect an element of circularity in the present characterisation of observability, due to the final clause, '. . . in making an observation '. If this is troublesome, substitute the pragmatic characterisation just discussed for this last occurrence of ' observation '. Thus the final clause would read '. . . in quickly deciding to accept . . . '. The considerations of the last section would then be interpreted as preliminary to the analysis of observability, the full analysis involving the instrumental considerations to be presently considered. This suggestion may sound a bit contrived, but I believe it to be the fairest way to interpret Maxwell and Carnap in order to avoid the abovementioned circularity.

[2] In ' Testability and Meaning ', Feigl and Brodbeck, *Readings in the Philosophy of Science*, p. 64.

Carnap would presumably maintain that ' is electrically charged ' is a non-observational predicate.

Another distinction that can be drawn is that between objects which *require high-powered instruments* to be observed and objects which can be observed with relatively low-powered instruments or with no instruments at all.[1] This is an important distinction (of degree) to draw, and one which *is* in fact made in ordinary circumstances. Let us see whether this distinction can serve the purposes of the partial interpretation theorists.

For many objects, this distinction will coincide with a continuum of size. Thus, we speak of the writing on the head of a pin as being invisible, thereby pointing out that it is too small to be seen with the naked eye, but requires a microscope. But small size is not the only reason why objects may require the aid of powerful instruments to be seen. For example, there are many stars which can be seen only with very high-powered telescopes. These stars would therefore have to be classified as unobservable objects, while stars visible to the naked eye would be classified as observable. But then is the term ' star ' an observation term or a non-observation term? We would have to put it in one category or the other, but neither *one* would be satisfactory. The alternative would be to distinguish between two terms, 'star_0 ' and ' star_u ' (say), where the former refers to stars visible to the naked eye, and the latter refers to stars visible only with a powerful telescope (unobservable stars). If we did this, a reconstruction of the ' theory of astronomy ' (or astrophysics) of the type advocated by the partial interpretation theorists would have to be such that the former term would be given a semantical rule whereas the latter would not, thus *assuming* that the *meanings* of these terms are of a radically different nature. I do not think that Carnap would be willing to accept this conclusion.[2] And yet it does follow if we simply characterise an unobservable object as one which can be observed only with high-powered instruments.

But Carnap could object that we have oversimplified the characterisation. We could, *in principle*, get closer to stars which we can only observe with a telescope from the earth, and thus see them without a telescope. Suppose, then, that we modify the characterisation to take into account what is *physically possible* (though perhaps not at present feasible). An observable object, then, would be one which can in principle be observed without the aid of any instrument. But this modification, I maintain, will still allow some objects as observable which Carnap and the authors in question would call unobservable. As an example, consider the predicate 'is electrically charged'. A physicist can determine, without the aid of any instrument, whether this predicate applies to some bodies. He need merely place his arm near a body in question and observe whether the hair on his arm stands on end. This would be an observation of the electric field even in Maxwell's sense; there would be no process of inference in deciding (quickly) whether the predicate should be applied. The physicist would merely place his arm near the body in question, and remark 'This body is electrically charged'. And yet Carnap has classified electric fields as unobservable entities.

---

[1] Not all those instruments which are ' complex ' are ' high-powered ', and conversely. (The latter notion itself involves several factors.) Also, a high-powered instrument may require very few preliminary adjustments; e.g. a good pair of binoculars.

[2] According to astrophysical theory, after all, the very same general things are true of both stars visible to the naked eye and stars observable only with a high-powered telescope.

Now it is true that the physicist has not determined the *exact intensity* of the field by observing it in this manner. For this, he *would* need some instrument. But he *can* determine in this way (roughly, to be sure) whether the field is weak or strong, and with practice he might be able to predict with some degree of accuracy what the reading of an instrument will be if placed near the body. He may even check whether his instrument is working properly in this manner! (Furthermore, relatively simple, 'low-powered' instruments are available which can be used to measure the intensity of some electric fields.)

The situation is quite similar, as a matter of fact, to what takes place when we observe the wind. We can tell whether the wind is strong or weak by merely looking out of the window, or by listening, or by putting an arm out of the window. An instrument would be *required* to determine the *exact wind velocity*, to be sure, but the wind would not be classified as unobservable *for this reason.*[1] Therefore, we can conclude that the characterisation under consideration, even as modified to take into account what is possible in principle, cannot be used to draw the distinction between observable and unobservable objects intended by the partial interpretation theorists. It cannot distinguish between winds and electric fields, both of which turn out to be observable entities; and Carnap has considered the various fields introduced in physical theories as clear examples of unobservable entities.

---

[1] If Carnap *would* classify winds as unobservable 'objects' *because* we require instruments to determine accurately their velocities, this would be extremely odd to say the very least. (Colours would be unobservable also, since we cannot determine their intensity 'accurately' without an instrument.) Furthermore, according to the partial interpretation thesis, a reconstruction of meteorology would have to be such that 'wind' (or its symbolic counterpart) is given no semantical rule, and we would have to understand an 'entire theory' before we could be said to understand the concept of a wind.

I must emphasise that I am not denying that these distinctions which can be drawn on the basis of various factors which arise from the use of instruments are valid or useful distinctions. Indeed, to be told that an object is unobservable, *meaning thereby* that it cannot be seen without a powerful microscope (for example), is to be told something interesting and important about the object. But I *do* deny that these distinctions will capture the sort of distinction *intended* by the authors in question. The examples they use will not be reflected by using the factors just considered as a basis for drawing the observable-unobservable distinction; and there is no implication that those terms which turn out to be non-observation terms *cannot* be given semantical rules when the theories in which they occur are reconstructed as interpreted logistic systems.

Finally, it must be emphasised that these distinctions are *different* distinctions, which have not been distinguished by the authors in question. They do not form part of some *general* distinction between observable and unobservable objects. If we refer to an object as unobservable, there are several *distinct* points that we may be making about the object. If we call electric fields *unobservable*, we may intend one of several things: we might be pointing out that instruments are *in fact* used in most cases to measure or detect them; or, we may be pointing out that instruments are *required* to determine their exact intensity in a reliable manner. On the other hand, we may call electric fields *observable*, thereby pointing out that they can sometimes be observed (or detected) *without instruments*. Similarly, we may refer to the writing on the head of a pin as *unobservable* (invisible) to point out that a microscope is *required* to see it; or, we may refer to it as *observable* to point out that with the aid of a microscope, it can be seen and read as easily as the print on this page. Thus even if we restrict ourselves to factors arising out of the use of instruments, there is no single thing that is always intended by referring to an object as observable or unobservable.

3. *Other Uses of 'Observable' and 'Unobservable'*

I shall now examine some uses of the 'observable-unobservable' terminology which have not been considered at all by the authors in question, but which are, nevertheless, of some importance.

There is a common use of the term 'observe' which is closely related to the concept of the *detection* of an entity. This is the use in which the positive outcome of a well-substantiated (or commonly used) test procedure for the presence of an entity will be described as an observation of the entity.[1] The positive outcome of the test is then viewed as 'the clincher', so to speak: 'Ah, two pulses of the appropriate type; we've finally detected the neutrino.' As such a test procedure becomes more commonly used, we might come to describe the observation of the pulses as an *observation of neutrinos.* The use of the cloud chamber for the detection of certain elementary particles has reached this stage. A physicist will speak of the observation of a particular type of track in a cloud chamber as *an observation of an electron.*

Events on the microphysical scale will also be spoken of as being *observed* in this manner. The observation of a certain type of forked track in a cloud chamber will be described as the observation of the creation of a positron-electron pair.[2] Physicists have even spoken of the observation of photons—entities which we are supposed to 'see with'. Max Born, in describing a certain (idealised) experiment, writes as follows:

> Now, according to the corpuscular view, the particle in the radiation process suffers a Compton recoil . . . the direction of which is undetermined to the same extent as is the direction in which the light quantum flies off after the process. Since the *light quantum is actually observed* in the microscope, this indeterminateness of direction is given by . . .[my emphasis][3]

I think it is clear that this usage of the word 'observe' will not provide a basis for drawing the distinction intended by the partial interpretation theorists. For it will turn out that there are almost *no* unobservable entities posited by well-confirmed physical theories.

---

[1] This will sometimes be described as the 'detection' of an entity, although the word 'detect' is often reserved for cases where the outcome was in doubt—in particular, where there was a question as to whether the entity existed in the first place, or where quite a bit of theoretical and experimental work immediately preceded the setting up of the apparatus which is used in the test for the presence of the entity in question.

[2] See Hilary Putnam, 'What Theories Are Not', in Nagel, Suppes, and Tarski (eds.), *Logic Methodology and Philosophy of Science* (Stanford University Press, 1962), p. 244.

[3] Max Born, *Atomic Physics* (5th edn.; New York: Haffner Publishing Co., 1950), p. 97.

Finally, there is a quite important use of the term 'observable' (actually, 'unobservable' is the term which carries the force in this distinction) which is intimately tied up with considerations based on the quantum theory. It is often said that electrons and the other elementary particles are 'unobservable in principle'. When a physicist says this, he intends to point out something quite specific about the structure of the laws of quantum theory in relation to some of the fundamental constants of nature which occur in these laws. It is still not clear what the proper interpretation of quantum theory should be, but one of the ways of describing what is meant by referring to electrons as 'unobservable in principle' is the following: We cannot construct an experiment which will allow us to determine simultaneously the precise position and momentum of an electron. The more precisely one of these is measured, the more uncertain the other will be. This is true because of the Heisenberg 'uncertainty relations'. Stated for position r and momentum p, the Heisenberg equation is

$$\Delta r \Delta p \geq h/2\pi$$

where h is Planck's constant, and Δr and Δp are the 'uncertainties' in the measured position and momentum, respectively, of the electron. If one of these quantities is very small, the other must be very large. This equation is sometimes made plausible as follows. In order to observe an electron, we must shoot light quanta at it, and detect the scattered quanta. If we wish to determine the *position* of the electron accurately, the quanta must be of a very small wavelength, hence a very high frequency. But according to the Planck-Einstein relation, $E = h\nu$, a light quantum of high frequency ν will also have a high energy E, so that the collision of the quantum with the electron will cause an 'unpredictable' change in the energy (hence the momentum) of the electron. Thus, to measure the position of the electron accurately is to spoil any chance of measuring its momentum accurately. The converse is also true. If we wish to measure the momentum of the electron accurately, we must use quanta of low energy so as not to disturb it; but quanta of low energy also have a low frequency, hence a large wavelength, therefore making a precise determination of the position of the electron impossible.

Now Planck's constant is quite small, but the mass of the electron is also quite small (the momentum p is simply the mass of the electron multiplied by its velocity, so that 'uncertainties' encountered 'really matter'). In other words, any precise measurement of the momentum will produce an uncertainty of position which is much greater than the *radius* of the electron; and conversely, a precise determination of the position will produce a change of momentum which is much greater than the *original momentum* of the electron. If, however, we are dealing with more massive particles, such as protons, light atoms, heavy atoms, large molecules, etc., the uncertainties tend to become negligible, although in theory they will always be present. Thus, in this sense of the term ' observable ', there is a continuum of size from entities such as baseballs, which are ' observable ', down to electrons, which are ' unobservable in principle '.

It is quite clear that the authors in question have not had *this* use of the ' observable-unobservable ' terminology in mind. It certainly will not serve their purpose. Regardless of what quantum field theory may evolve, the partial interpretation theorists have wanted to call certain entities unobservable which do not fall into the province of elementary quantum theory, e.g. various fields. Also, it is as if ' unobservable-in-principle ' came in a unit, as I have just written it, rather than ' in principle ' being added to a prior sense of ' unobservable '. To call an electron unobservable-in-principle is a shorthand way of saying something quite specific about the structure of quantum theory in relation to the mass of the electron. Furthermore, the continuum I mentioned above is a quite precisely defined continuum; it is not just that there is a ' fuzzy border area ' between observable and unobservable entities. Therefore it would be unacceptable to draw a line ' for simplicity ', and proceed to treat the two categories of entities in distinct ways, as is done in the partial interpretation analysis.

4. *Conclusions*

We may summarise the conclusions reached so far as follows:

The specific analyses of the notion of an observation term (or observable entity) offered by the defenders of the partial interpretation thesis are unacceptable. To the extent that they support a distinction between two types of entities at all, the distinction will not be of the type they require in order to maintain the partial interpretation thesis.

The further suggestions which they have made concerning the use of instruments as a basis for drawing a distinction between observable and unobservable entities will not serve their purpose either. We saw that there are several distinctions that could be drawn with respect to the use of instruments.

Finally, a consideration of some of the ordinary uses of the ' observable-unobservable ' terminology (including ordinary scientific uses) showed that there are in fact *several* uses of this terminology, none of which lend support to the partial interpretation thesis. If we refer to an object as observable (or unobservable), there are several points that we may be making about the object. Which of these is intended will depend on the context in which the assertion ' Xs are observable ' (or ' unobservable ') is made. There is no single distinction which the ' observable-unobservable ' label covers which is the same in all contexts in which it is actually used—ordinary or scientific. If we hear the sentence ' Xs are unobservable ' out of context, we still do not know what exactly is being said about Xs. We would have to know who has said it to whom for what purpose; and depending on these latter factors, the assertion may be true or false, even for one and the same entity. To put the point bluntly, there is no such thing as *an observation term* or *an observational vocabulary* (period !).

(*to be concluded*)

A LOGIC FOR EVIDENTIAL SUPPORT*

L. JONATHAN COHEN

ACCORDING to a well-known conception of inductive argument we do well to test our hypotheses in a variety of ways in order to make as sure as we can that no other factors are relevant to our problems than the ones we suppose. Bacon's tables of presence and absence, Hume's recommendation to ' enquire by new experiments if every particular circumstance of the first experiment was essential to it ', Whewell's emphasis on the consilience of inductions, Mill's canons of agreement and difference—all these in their different ways were tributes to the importance of eliminative induction. Of course, the older philosophers for the most part conceived the conclusions of such inductive arguments in a purely qualitative fashion. The validity of a hypothesis was thought of as being either conclusively confirmed or conclusively disconfirmed or left an open question. But comparative, and perhaps even quantitative, conclusions are by no means obviously impossible, and the object of the present paper will be to explore the logic of a confirmation-functor c[H, E] denoting the degree of support by eliminative induction that one statement, E, gives to another, H.

I shall begin by laying down informally certain conditions to which this functor might reasonably be supposed to conform, and I shall show that it must therefore conform to certain other conditions, which I shall also claim to be reasonable. In particular it turns out to conform to conditions which ensure that it cannot denote any function of any relevant probabilities. I shall then develop a generalised modal logic within which the functor is definable in terms of the primitive symbols and its logical properties are formally provable. I shall not claim that the concept thus explicated is the only possible concept of degree of support by eliminative induction, still less that it is the only possible concept of degree of evidential support. But in the course of my argument I shall give some examples to suggest that the concept has not been altogether without implicit use in the actual advancement of human knowledge.

* Received 14.ix.65

I

Let us assume, first, that logically equivalent statements may be substituted for one another in our confirmation-functors. I shall term this the Equivalence Condition. It asserts that

(1) For all statements E, E′, H and H′, if H is logically equivalent to H′, and E to E′, then c[H, E] = c[H′, E′].

Now, no confirmation-theorist that I know has rejected the equivalence condition for evidential statements, viz. the requirement that if E is logically equivalent to E′ then c[H, E] = c[H, E′]. But it is sometimes suggested that in order to solve Hempel's paradoxes of confirmation we should reject the equivalence condition for hypotheses, viz. the requirement that, if H is logically equivalent to H′ then c[H, E] = c[H′, E]. For, if we reject this condition, we need not suppose—as it is indeed paradoxical to suppose—that the report of a non-black non-raven provides as good evidential support for ' All ravens are black ' as for its contrapositive equivalent ' All things that are not black are not ravens '. But such a solution of Hempel's paradoxes is bought at a very high cost. If we suppose that confirmation-functors of the form c[H, E] are resistant to substitution of logical equivalents in their first argument-place, we assign that place to a deeper level of non-extensionality—to a level of referential opacity that is murkier—than any to which we need assign any argument-place in an assertion that such-or-such a statement is a law of natural science, mathematics or formal logic. What typically belong on this level are reports about what people have actually said or about the content of their minds. We certainly cannot be sure that a man who says, or entertains the thought, that H is true also says, or entertains the thought, that H′ is true, even if H and H′ are logically equivalent to one another. But the logical liaisons of confirmation-functors will be very restricted indeed if we reject the equivalence condition for hypotheses. For we must then also reject or restrict the Consequence Condition, viz.

(2) For all E, H and H′, if H is logically deducible from H′ then c[H, E] $\geqslant$ c[H′ E],

since the equivalence condition for hypotheses is a logical consequence of (2).

But how far can we afford to restrict (2)? Certainly we should have to retain the thesis that evidential support is passed on by a universal statement to its substitution-instances. There would be little point in scientific generalisation if the applications of a universal hypothesis H to particular cases were not at least as well supported as H itself by the evidence for H: our general knowledge would be useless. But just the same would be true of scientific system-building—the axiomatisation of a body of universal hypotheses—if the Consequence Condition (2) were not accepted in its usual, unrestricted form. There would be little point in such systematisations if evidential support were not automatically passed on by a conjunction of universal statements to any other universal statement logically deducible from it. I conclude that the Consequence Condition (2) cannot be safely restricted in such a way as to invalidate the equivalence condition for hypotheses, and that any quantitative or comparative concept of evidential support that failed to satisfy the latter condition would be of relatively little use or importance. Hempel's paradoxes of confirmation are therefore better solved in some other way than by the rejection of (1)—and an alternative solution will in fact emerge later from the retention of (1): see 6 below.

The second condition on evidential support that I want to assume is what I shall term the Instantial Comparability Condition. According to this assumption not only do inequalities in degree of evidential support between two theories lead normally to inequalities in degree of evidential support for the applications of these theories to individual cases, but also inequalities in the latter normally stem from inequalities in the former. More precisely

(3) For all E, E′, S, S′, U and U′, if
  (i) U is a universal statement, S is just a substitution-instance of H, and E is a conjunction of statements each of which predicates something of some specified element in the domain of U, other than the element referred to by S, and
  (ii) U′ is a universal statement, S′ is just a substitution-instance of U′, and E′ is a conjunction of statements each of which predicates something of some specified element in the domain of U′ other than the element referred to by S′,

then $c[U, E] > c[U', E']$ if and only if $c[S, E] > c[S', E']$. (By 'just a substitution-instance' here I mean that all the relevant information given by S and S′ is that the elements to which they refer satisfy U and U′, respectively: their modes of referring to these elements add no further relevant information.)

If we come to make new tests on the basis of which we conclude a scientific theory to be even better supported by evidence than we had hitherto supposed, we would normally infer a correspondingly increased measure of support for each singular prediction or retrodiction we derive from that theory. Of course, if the fact described by S′ were included in the evidence for U′, c[S, E] might well be thought not greater than c[S′, E′] even though c[U, E] was thought greater than c[U′, E′], since S′ would then be a logical consequence of E′. But clause (ii) of the Instantial Comparability Condition excludes such a situation by requiring that E′ be a conjunction of statements each of which predicates something of some specified element in the domain of U′ other than the element referred to by S′. Similarly, if a man is told that a singular prediction—about, say, the curative value of the pills in *this* box for *his* illness—is better supported by the evidence now available than by the evidence previously available, he would normally suppose that the evidence now available lends greater support to some general hypothesis about the curative value of *that kind* of pill for *that kind* of illness. The concept of evidential support with which the Instantial Comparability Condition is concerned is thus one that differs, broadly, from Popper's main field of interest in that it is as concerned with the extent of support for singular predictions as with that for general hypotheses. It also differs, broadly, from Carnap's main field of interest, in that it relates all questions about the extent of support for singular predictions to questions about support for general hypotheses which these predictions instantiate. It presupposes the uniformity of nature in regard to evidential comparisons. Indeed it is difficult to see how in eliminative induction we can avoid concern with general hypotheses. The reason why we suppose it useful to vary the circumstances in which we test a hypothesis may be that in this way we can eliminate, by falsifying, certain alternative general hypotheses. Or it may be that in this way we can extend the domain to which our own hypothesis is not excluded from applying. But in either case we shall only pay attention to those features of an individual situation in virtue of which it succeeds or fails in instantiating some general type of situation.

I shall now show that any concept of evidential support that satisfies the Equivalence and Instantial Comparability Conditions must also satisfy what I shall term the Instantial Conjunction Conditions, viz.

(4) For all E, S^1, S^2, . . . S^n and U, if U is a universal hypothesis and S^1, S^2, . . . S^n are just some substitution-instances of U, and

> E is a conjunction of statements each of which predicates something of some specified element in the domain of U other than any of the elements referred to by S^1, S^2, ... S^n, respectively, then $c[S^1, E] = c[S^1 \& S^2 \& \ldots \& S^n, E]$.

If U^1 is a universal hypothesis with a domain of at least *n* elements, there are other hypotheses U^2, U^3 ... U^n, logically equivalent to U^1, formed by generalising analogously not about the individual elements of U^1's domain of discourse, but about pairs, trios ... or *n*-membered classes of these elements. (At any rate this is true if the universal hypothesis U^1 is not taken to be a generalised truth-functional conditional: if it is so regarded then U^1 must be taken as the conjunction of this conditional with the statement that at least *n* individuals satisfy its antecedent, and U^2, U^3, ... U^n must each be taken as the conjunction of a truth-functional conditional about *i*-membered classes, where $i = 2$, or $i = 3$, or.... $i = n$, as the case may be, with the statement that at least one such class satisfies this conditional's antecedent.) If U^i is such an equivalent of U^1, the Equivalence Condition (1) gives us, for any E,

$$c[U^1, E] = c[U^i, E]. \tag{5}$$

Also for any evidential statement that element^1, element^2, ... and element^n of U^1's domain have certain features there is an equivalent evidential statement that the class which has all and only these elements as its members is a class of entities that severally have these features. Let us suppose that E^i is the conjunction of all such equivalents for statements logically implied by E^1, where *i* is the number of elements in each of the classes that form the domain of U^i. That would give us from (5), by the Equivalence Condition again,

$$c[U^1, E^1] = c[U^i, E^i]. \tag{6}$$

For example 'All pairs of ravens are pairs of black things' has the same degree of support from some suitable E^2 as has 'All ravens are black' from E^1.

Now it is a logical consequence of the Instantial Comparability Condition (3) that in the circumstances described by clauses (i) and (ii) of that condition $c[U, E] = c[U', E']$ if and only if $c[S, E] = c[S', E']$. Hence, more specifically, if in these circumstances we are given (6), we may derive

$$c[S^1, E^1] = c[S^i, E^i]. \tag{7}$$

For example, E^1 supports the prediction that if *a* is a raven *a* is black to just the same extent as E^2 supports the prediction that if the pair *a*, *b* is a pair of ravens it is a pair of black things. Let $S^1, S^2, \ldots S^{i'}$ be the predictions about individuals that correspond to S^i in the way that, e.g. ‘ If *a* is a raven *a* is black ’ and ‘ If *b* is a raven *b* is black ’ correspond to ‘ If the pair *a*, *b* is a pair of ravens it is a pair of black things ’. Then by the Equivalence Condition we may obtain

$$(8) \qquad c[S^i, E^i] = c[S^1 \mathbin{\&} S^2 \mathbin{\&} \ldots \mathbin{\&} S^{i'}, E^i]$$

and from (7) and (8) we obtain, with a further use of the Equivalence Condition

$$(9) \qquad c[S^1, E^1] = c[S^1 \mathbin{\&} S^2 \mathbin{\&} \ldots \mathbin{\&} S^{i'}, E^1].$$

And thus we have proved (4), the Instantial Conjunction Condition, on the assumption of the Equivalence and Instantial Comparability Conditions, (1) and (3).

We may, however, note in passing that the Instantial Comparability and Conjunction Conditions are only of permanent interest in relation to universal hypotheses with infinite domains. For if the domain of U were finite, there might come a time when we had so much evidence available about the elements of that domain that, if we were to cite all available evidence, in accordance with the usual rule (cf. 2 below), we could no longer find a substitution-instance, S, of U satisfying the requirement that S does not mention any element of U's domain which is mentioned in E.

2

I have proved elsewhere[1] that, so far as a confirmation-functor is to satisfy the equation $c[H, E] = c[H \mathbin{\&} H', E]$, as in the consequent of (4), it cannot, except in certain limiting cases, even denote a function of the relevant probabilities—p[H, E], p[E, H], p[H] and p[E]—let alone a simple probability. At least in the circumstances specified in (4), therefore, no confirmation-functor with a syntax articulatable in the probability calculus can serve to measure or compass degrees of evidential support in a way that satisfies (1) and (3). But the circumstances specified in (4) restrict the validity of this result to estimates of the evidential support for substitution-instances of generalisations. I shall now argue, at least in the case of universal hypotheses, in favour of a General Conjunction Condition which would give the same anti-probabilistic results. The General Conjunction Condition asserts that

---

[1] In ‘ What has Confirmation to do with Probabilities? ’ in *Mind* lxxv, Oct. 1966. In that article an earlier version of the proof of the Instantial Conjunction Condition occurs. This conclusion, or rather a closely similar one, is proved there not from the Instantial Comparability Condition but from what I called the Predictive Comparability Condition. I have now abandoned that form of the argument partly because it is more complex than the present one, and partly because it treats statements of initial conditions on a par with statement of evidence. The present paper formed the substance of two lectures delivered on 26th and 27th March 1965 at the University of Liège and the Centre National Belge de Recherches de Logique, respectively. I am indebted to the Editor, Dr Mary Hesse, for several very helpful comments and suggestions.

(10) For all E, H and H′, if $c[H', E] > c[H, E]$, then $c[H \& H', E] = c[H, E]$.

The kind of argument that was used to establish (4) can take us a little bit further, but will not serve to establish (10). We can certainly prove (9) where $S^1, S^2, \ldots S^i$ are each a substitution-instance of a different generalisation, so long as each of these generalisations has the same degree of support from the evidence as its conjunction with the others. For example, if ‘ All ravens are black ’ has the same degree of support from E as has ‘ All ravens are black and all swans are white ’, then it also has the same as ‘ All ordered pairs of one raven and one swan are ordered pairs of one black thing and one white thing ’ (with the above qualification about truth-functional analyses); and from this it follows that the prediction ‘ if *a* is a raven *a* is black ’ has the same degree of support from E as has ‘ If the ordered pair *a*, *b* is an ordered pair of one raven and one swan it is also an ordered pair of one black thing and one white thing ’, and thus the same as ‘ If *a* is a raven, *a* is black, and if *b* is a swan, *b* is white ’, so long as E says nothing about *a* or *b*. In addition, we can prove that one generalisation has the same degree of support from the evidence as has its conjunction with certain others, so long as all the generalisations concerned are substitution-instances of some single higher-order generalisation and the evidence does not relate specifically to one of the generalisations any more than to each of the others. For then the proof goes through exactly as for singular predictions that are substitution-instances of the same generalisation. The Instantial Comparability Condition applies to the derivation of first-order generalisations from second-order ones in very much the same way as it applies to the derivation of singular predictions from first-order generalisations. For example, ‘ If lions are vertebrates, they have kidneys ’ and ‘ If tigers are vertebrates, they have kidneys ' are both derivable by instantiation from ' If any kind of animal is a kind of vertebrate, it is also a kidney-possessing kind ', and each of the derived generalisations has the same degree of support from appropriate evidence as has their conjunction.

Arguments of this nature, however, do not seem to take us any further. Yet it is obviously unsatisfactory to have to work with two syntactically different confirmation-functors, a non-probabilistic one applicable in the cases so far considered and a probabilistic one applicable in a vast number of other cases. I shall therefore now propose two further conditions, both of which seem reasonable assumptions for a logic of eliminative induction, and I shall show that (10) is demonstrable from these, with the aid of (2), at least in the case of universal hypotheses within the same general field of enquiry.

In the first place we must allow for the fact that in eliminative induction one unmistakable counter-instance suffices to justify eliminating a universal hypothesis, at least in the form in which that hypothesis has thereby been falsified. (By a ' counter-instance ' to a universal hypothesis U I mean a member of U's domain that lacks the characteristic ascribed by U to all members of its domain, e.g. the characteristic of being a raven only if black, where U is ' All ravens are black '.) Of course standards of unmistakableness vary. But even if it is required that the falsifying experiment should have been repeated a certain minimum number of times, *m*, the same point can still be made here. We merely have to consider instead the logically equivalent hypothesis that generalises analogously about *m*-membered classes of individuals. Again, it is quite true that in many cases the hypothesis may only need to be modified or qualified in some respect in order to escape this kind of falsification. Nevertheless in its unmodified form it must be rejected, and therefore has zero degree of evidential support if the available evidence includes an unmistakable counter-instance. Contrapositively, if a hypothesis has some positive, non-zero degree of support from certain evidence, that evidence cannot include a counter-instance and can therefore give no support—at least no direct support by eliminative induction—to the hypothesis's negation. For in measuring positive support here we are out to weigh just the accumulation of varied unsuccessful attempts to falsify a hypothesis, and if the evidence includes no counter-instance to a hypothesis it includes no unsuccessful attempt to falsify that hypothesis's negation. Thus we have, for universally or existentially general hypotheses, what I shall term the Negation Condition, viz.

(11) If $c[H, E] > 0$, then $c[-H, E] = 0$.

But the converse of (11) is not to be assumed. If we have $c[-H, E] = 0$ it does not follow that we must accept $c[H, E] > 0$. Admittedly, if we have $c[-H, E] = 0$ because E falsifies $-H$, then we may have to accept $c[H, E] > 0$. But E may not describe any situation of a kind that, according to our confirmation-metric, is important for the confirmation of H, or E may give zero support to $-H$ because it is totally irrelevant to both H and $-H$, and in either case we must accept $c[H, E] = 0$.

Secondly, we must no doubt always use all relevant available evidence in assessing the merit of our hypotheses, whatever confirmation-theory we adopt.[1] More specifically, in eliminative induction, if the first argument of a confirmation-functor is a universal hypothesis U, then the second, E, whatever else it states, must at least implicitly either report at least one counter-instance to U or deny that any is included in the available evidence. Otherwise such a functor is useless and we must regard the evidence as irrelevant. For if E could not be relied on to report at least one counter-instance, if any be available, then, even given the truth of E, we could not reasonably adjust our confidence in U, on E, to whatever favourable evidence is included in E, since none of this evidence would have any confirmatory value in eliminative induction if a single counter-instance were available. But just what should be the criteria of availability in the case of evidence for eliminative induction? Clearly it is not sufficient to go merely on all the evidence that happens to be at hand, since the severer tests on a hypothesis may have been deliberately omitted. One must require instead that each hypothesis to be assessed must have been submitted, if possible, to all tests normally thought practicable and worthwhile in its particular field of enquiry. I say ' if possible ' here, because in some cases there may not have been time to conclude all these tests, and one does not want to restrict unduly the times at which assessments of evidential support may be uttered. But to guard against such a premature assessment's being any less reliable than it need be we can, and should, also place a condition on the order in which the tests may be assumed to have been carried out. We must assume not only (i) that, whatever else it states, E is irrelevant to U unless

---

[1] Cf. R. Carnap's Requirement of Total Evidence in *Logical Foundations of Probability*, 1950, p. 211 and K. R. Popper's insistence that his corroboration functors apply only to hypotheses that have been tested with maximum severity, *Logic of Scientific Discovery*, 1959, p. 418.

(at least implicitly) E either reports at least one counter-instance to U or to another universal statement U′ from which U is logically deducible or denies that a counter-instance to U has emerged from such-and-such tests, the results of which constitute all the available evidence, or denies that a counter-instance to U′ has so emerged, and (ii) that all tests normally thought worthwhile in each field of enquiry are carried out on each universal hypothesis to be assessed, but also (iii) that the tests are carried out in a fixed serial order, such that where one test is thought severer—more likely to falsify—than another the former precedes the latter. I shall term this whole requirement the Evidential Relevance Condition, and I suppose that the mark of one test's being thought severer than another is that success in escaping falsification by the former scores more for a hypothesis than success in escaping falsification by the latter. But the Evidential Relevance Condition is not, of course, directly a condition on the logical syntax of confirmation-functors, as are (1), (2), (3), (4), (10), and (11). In itself it is just a partly semantical, partly pragmatic condition on the values and arguments of confirmation-functors, when these are assessing degree of evidential support by eliminative induction.

It is easy to see that any confirmation-functor obeying the Consequence, Negation and Evidential Relevance Conditions must obey the General Conjunction Condition (10), at least so far as universal hypotheses within the same field of enquiry are concerned. There are three cases to be considered:

Case i: $c[H', E] > c[H, E] = 0$. Here (10) follows in virtue of the Consequence Condition (2).

Case ii: $c[H', E] = c[H, E] > 0$. Here it follows from (11) that E reports no counter-instances to H or H′. Therefore according to the Evidential Relevance Condition, however much a hypothesis scores for passing a particular test successfully, H and H′ must have passed the same tests, because they have the same scores. But, if H and H′ have both passed certain tests by eliminative induction, they are both unaffected by the variations of circumstance involved and also therefore any universal hypothesis logically equivalent to H & H′ is unaffected. Hence H & H′ must have the same score as H or H′ has.

Case iii: $c[H', E] > c[H, E] > 0$. Here too it follows from (11) that E reports no counter-instances to H or H′. Hence by the Evidential Relevance Condition we know that any universal hypothesis logically equivalent to H & H′ has passed all the tests that H has, though no more, and therefore H & H′ has the same score as H.

3

So far I have only proved the General Conjunction Condition (10) for universal hypotheses within the same field of enquiry. Of course, it can easily be shown that (10) embraces the Instantial Conjunction Condition (4) as a special case, at least where $n = 2$. For in the circumstances specified in (4) it follows logically from the Instantial Comparability Condition (3) that $c[S^1, E] = c[S^2, E]$, and from this the consequent of (4) follows by detachment from (10). But I shall not attempt similar pre-systematic demonstrations of (10) for every variety of statement involved in eliminative induction. Instead I shall now assume that, in any formal systematisation of the logic of these confirmation-functors for eliminative induction, it is legitimate to include (10) as a thesis. It seems a legitimate simplification to suppose that a conjunction-condition that is demonstrably applicable in such standard cases is also applicable elsewhere. Accordingly, after considering several possible objections to (10), I shall propose a formal logic for confirmation-functors within which (10) is derivable.

Most resistance to (10) is likely to come from those who hold, probabilistically, that if $c[H', E] > c[H, E]$ then $c[H \& H', E] < c[H, E]$. However at least one argument of an opposite nature is worth considering.

Mario Bunge[1] has held, contrary to the Consequence Condition (2), that if two conjuncts are unequally confirmed their conjunction is better confirmed than the less well confirmed conjunct. His argument is that in actual practice no theory is rejected solely because one of its assumptions is suspected or even found to be partially false. ' Even if no adequate substitute for the partially false assumption is found ', he says, ' the theory may still be of some use and will be said to contain a grain of truth.' But this argument would support my thesis, not Bunge's. If one element of a conjunction is only partially false, and and the other entirely true, we might well regard the whole conjunction as only partially false and therefore also as partially true. This is a familiar thesis in multi-valued logic: the truth-value of a conjunction is the lesser of the two truth-values of its conjuncts. The trouble is that Bunge's axiomatised theory of confirmation—his syntax of our ' estimate of the material adequacy of factual propositions '—goes considerably beyond this. The theory entails, for instance, that if

[1] *The Myth of Simplicity*, 1963, pp. 118 ff.

one of two conjuncts is totally false and the other not, then the conjunction has a higher value than total falsehood (theorem 11). Admittedly there would be an ‘ element of truth ’, as Bunge puts it, even in this conjunction, and in a sense therefore the conjunction would be said to be partially true. But being partially true in the sense of containing one true element is not the same as being partially true in the sense of having as a whole some truth-value between truth and falsity. The fact is that if we know one of the conjuncts to be false we have all the evidence we need, as a matter of logic, for the conjunction's not being worth accepting as a whole, even though, as a matter of methodology, we may sometimes in practice decide to accept if for the moment because we have nothing better to put in its place. I conclude therefore that Bunge has not produced a valid argument for supposing a conjunction is ever better supported by given evidence than the least well supported of the conjuncts that compose it.

4

A very different objection to (10), the suggestion of which I owe to correspondence with Professor J. L. Mackie, might run like this. ‘ If $c[H, E] = c[H', E]$, so too $c[-H, E] = c[-H', E]$, and so too, therefore, by (10) $c[-H, E] = c[-H \& -H', E]$, and so again $c[-H, E] = c[-(-H \& -H'), E]$, and so finally $c[H, E] = c[HvH', E]$. But it is easy to think of E, H and H′ for which this last equation would be grossly paradoxical, even though $c[H, E] = c[H', E]$. For example, where H is “ Drug addiction tends to promote mental instability ”, H′ is “ Mental instability tends to promote drug addiction ”, and E reports the evidence of correlations between mental instability and drug addiction. There is, of course, no firm ground here for objection to the Instantial Conjunction Condition (4): perhaps $c[HvH', E]$ is never greater than $c[H, E]$ where H and H′ are both substitution-instances of a single generalisation and neither refers to any element described in E. But the General Conjunction Condition (10) is clearly untenable.’

I too would want to reject the thesis that if we have $c[H, E] = c[H', E]$ then we must accept $c[H, E] = c[HvH', E]$. All that can safely be accepted in such circumstances, about degree of evidential support for disjunctions, is what follows from the Consequence Condition (2), viz. $c[H, E] < c[HvH', E]$. But by rejecting the thesis that c[H, E] = c[H′, E] implies c[H, E] = c[HvH′, E], I am not compelled to reject the General Conjunction Condition (10). The former thesis is not demonstrable from (10) alone. It is demonstrable from (10) only in conjunction with a double use of the assumption

(12) For all E, H and H′, if c[H, E] = c[H′, E] then c[−H, E] = c[−H′, E].

That is, the objection under consideration assumes that if any two statements are equally supported by given evidence so too are their negations. But why should we accept this assumption? Certainly we should have to accept (12) if we held, like Carnap, that confirmation-functors have exactly the same syntax as probability-functors, since we should then accept, for all E and H, c[−H, E] = 1−c[H, E]. We should also have to accept (12) if we held that degrees of confirmation were certain sorts of functions of probabilities. For example, we might then have to accept, as apparently Popper does,[1] c[−H, E] = 0−c[H, E]. But, given the Negation Condition (11) we are only forced to accept (12) where we have c[H, E] = c[H′, E] > 0. Where instead we have c[H, E] = c[H′, E] = 0, we are not forced to accept (12). For E may fail to support H through falsifying it, so that we have c[−H, E] > 0, while E fails to support H′ through irrelevance, so that we have c[−H, E] = 0. Now the objection under consideration first uses (12) where c[H, E] > 0 is given—which is all right. But it then seeks to use (12)—putting −H for H—where we have to accept, according to (11), c[H, E] = 0. And that step is invalid.

Perhaps part of the temptation to accept (12) even where we have c[H, E] = 0 springs from an assumption that c[H, E] = 0 implies c[−H, E] = 1. But this is a probabilistic assumption which has no place in the logic of eliminative induction as envisaged here. Even when −H is deducible from E we do not have to give c[−H, E] its maximum value. We are tempted, because c[H, E] = 0 when E logically implies −H, to believe that therefore c[−H, E] = 1 then. This belief rests on the conception of evidential support for generalisation as something of which logical implication is a limiting case. But thus to treat deduction as a limiting case of eliminative induction is radically to blur the nature of the latter.

Consider what is presupposed by so treating deduction. The laws of formal logic determine whether our attempts at deductive argument are valid or not. Hence to suppose, in confirmation-theory, that deduction is a limiting case of induction is to suppose that what a confirmation-metric should supply for induction is a set of gradings for the validity of arguments from evidence to hypotheses, as if in inductive reasoning we were trying to make fully valid deductions but most of the time were neither completely succeeding nor completely failing. It is as if the object of a confirmation-metric were to provide a method of scoring the extent of the success or failure of a natural scientist to behave like a pure mathematician. But if natural scientists were really just trying to behave like pure mathematicians, and deduce formally valid conclusions from reported evidence, they could get no further with a universal hypothesis than establishing whether or not it had yet been refuted. There could be no difference for them between one unrefuted universal hypothesis and another. The evaluation of such differences, in the light of the available evidence, would be an entirely unimportant activity. Yet, unless we think this activity an important one, confirmation-theory itself is unimportant. Hence we must suppose that, in evaluating the difference between one unrefuted general hypothesis and another, natural scientists are not just attempting to carry out deductions in a kind of limbo between formal validity and formal contradiction but instead are applying autonomous principles for the measurement of evidential support. It is the task of confirmation-theory, as I conceive it, to elucidate the nature of these principles and to distinguish clearly their field of application from the field to which the laws of formal logic are applied. I therefore conceive it as a merit, not a fault, in a theory of evidential support if its syntax does not require us to accept $c[-H, E] = 1$ where $-H$ is logically deducible from E.

---

[1] *The Logic of Scientific Discovery*, 1959, pp. 400 f.

To put it another way, the domain of a confirmation-metric is not the interval between E's contradicting H and E's implying H, but rather the interval between E's implying there to be no support at all for H, on available evidence, and *E*'s implying *H* to state a physical necessity or law of nature. It is the latter result that constitutes the kind of success at which eliminative induction aims. And, even if E contradicts H, it certainly does not follow that E must therefore imply $-H$ to be a law of nature and so that $c[-H, E]$ must have its maximum value.

5

A third—more familiar—kind of objection to my General Conjunction Condition (10) might run as follows. 'The evidence that both of two evenly balanced six-sided dice, one white and one black, have fallen with an odd number uppermost gives a 1/9 chance of truth to the hypothesis that both have fallen with five spots uppermost, a 1/3 chance to the hypothesis that the white die has fallen thus, and a 1/3 chance to the hypothesis that the black die has done so. But this is a typical example of the way in which, contrary to (10), the same evidence can support a lower degree of confirmation for the conjunction of two statements than for either statement separately. Hence (10) must be rejected.'

My answer to this kind of objection is to deny that I am concerned at all with evidential support in this sense. What distinctively characterises the situation here is that the statement of so-called evidence logically implies the statements about the chance certain hypotheses have of being true. Accordingly, if we say here that $c[H, E] = c[H', E] = 1/3$ and $c[H \& H', E] = 1/9$, we are in effect accepting that equations of the form $c[H, E] = n$, where $1 > n > 0$, may be logically true. Now for Carnap this is quite permissible. For him *all* equations of the form $c[H, E] = n$ are as *a priori* as truth-functional tautologies, and this is quite in keeping with the overall supposition that deduction is a limiting case of induction. (Compare his thesis[1] that a statement of deductive logic, like ' E L-implies H ', means that the entire range of E is included in that of H, and can be established by a logical analysis of the meanings of E and H provided that a definition of ' L-implication ' is given, while a statement of inductive logic, like $c[H, E] = 3/4$, means that three-fourths of the range of E is included in that of H, and can be established by a similar logical analysis, provided that a definition of ' degree of confirmation ' is given.) But there is another concept of evidential support in accordance with which it is impermissible to treat most equations of the form $c[H, E] = n$ as being capable of logical, *a priori* truth; and that is the concept, not Carnap's, which I am concerned to elucidate.

Indeed, the replacement or revision of procedures for assessing the acceptability of hypotheses of a certain kind on given evidence is quite a familiar event, so long as we do not confine our view to examples of the kind produced by Carnap to illustrate his theory (e.g. the chess tournament).[2] Suppose a hypothesis about the non-toxicity of some pharmaceutical preparation, such as thalidomide, comes out sufficiently well under the established battery of tests to

[1] *Logical Foundations of Probability*, pp. 201 f.

[2] *Logical Foundations of Probability*, pp. 382 f. 394 ff.

justify accepting it, but nevertheless a number of poisoning cases later on show that the hypothesis was false. Obviously the battery of tests needs revision. Certain principles have been shown empirically to be unreliable, viz. the principles in accordance with which evidential support for the hypothesis about thalidomide was previously assessed. Admittedly the logical syntax of our confirmation-functors may impose some *a priori* restrictions on the measures of support certain kinds of evidential statements may lend to certain kinds of hypotheses. For example, there are the restrictions imposed by the Equivalence, Consequence, Instantial Comparability, Conjunction and Negation Conditions. But apart from restrictions imposed by syntactical laws such as these, the assignment of values to c[H, E] for any particular kind of H and E must not be taken as a logical and wholly *a priori* problem, if confirmation-theory is to grapple at all with the realities of eliminative induction and not just shadow-box with artificial languages that are too simple and jejune to be useful for any practical purpose. The logical, relatively *a priori* problem about confirmation-functors, as I regard them, is solely to determine their syntax. Within the overall framework of that syntax theories about the proper confirmation-metric for hypotheses in this or that field of research—theories assigning certain scores to hypotheses that have passed certain tests and postulating a certain minimum score, to be called here a ' threshold of reliability ', as a sufficient condition for the justifiable acceptance of such a hypothesis as a guide to action—are in principle corrigible under just the same empirically detectable circumstances as are the hypotheses themselves. If a metric is not thus corrigible, no argument based on it can touch the validity of the General Conjunction Condition (10) within its intended field of application. Now no observable event would count against the view that, if two evenly balanced six-sided dice, one white and one black, have fallen with an odd number uppermost, this is adequate ground for ascribing a 1/9 chance of truth to the hypothesis that both have fallen five spots uppermost, 1/3 chance to the hypothesis that the white die has fallen thus, and a 1/3 chance to the hypothesis that the black die has done so. Hence these ascriptions, in the sense in which they are true, cannot justify the rejection of (10).

If we are to envisage any relevant confirmation-metric as being empirically corrigible, we should presumably envisage a possible plurality of confirmation-metrics, each applicable to hypotheses about a certain kind of subject-matter, rather than some single, comprehensive metric applicable to all hypotheses that we can formulate. The Carnapian quest for a single, all-embracing confirmation-metric seems not unreasonable if the metric is supposed to have *a priori* validity and is regarded as a branch of logic. But if instead a confirmation-metric is regarded as just another hypothesis or theory of empirical science there is no more reason to suppose that we ought in any stated period of time to be able to attain a single, all-embracing confirmation-metric than that we ought to be able to attain a single all-embracing theory in the rest of science. Indeed confirmation-metrics of the kind I envisage would be much more useful in assessing the hypotheses of relatively primitive sciences, such as those of pharmacology or psychology, than in dealing with theoretical physics or other advanced sciences. In an advanced science the range of factors bearing on experimental situations is relatively well understood, and the laws controlling them are well-known. The scope of a new theory can therefore normally be tested by experiment so determinately that there is little need for a confirmation-metric. The need for a confirmation-metric is most seriously felt when, as in a primitive science, we do not feel confident that we fully understand all the relevant factors and so the outcome of our experiment is by no means fully determinate. We need then to measure the extent to which anything at all has been established. And just as the hypotheses of a primitive science are characteristically not yet unified by some single, all-embracing model or theory, so too we may expect to find the relevant confirmation-metrics relatively ununified. In pharmacology, for example, a metric that fits pre-clinical trials of antibiotics may conceivably be quite inappropriate for pre-clinical trials of vaccines.

I do not want to claim that some metrics of this kind are already in explicit use, or even that they ever will be. Philosophical idealisations are not necessarily the best instruments of day-to-day science. But it is apparent from text-books on experimental chemotherapy, for example, that comparisons are sometimes drawn which come very close to implicitly invoking such a metric. We are told that pre-clinical trials for toxicity are more reliable if conducted on two species of animal than on one, more reliable again if conducted on one rodent species and one non-rodent than just on two rodent species or two non-rodent, more reliable still if conducted on three species than on two, and so on.[1] But how much more reliable, we may want to ask, in each case? How does the increase in reliability achieved by going from one species to two compare with the increase achieved by going from two to three? Of course, within the tests on any one species we are able to determine mathematically the consequences of working with samples of such-or-such a size. For example, if groups of ten animals are used for each dose of the drug we may determine the median lethal dose for a particular species with 95 per cent confidence limits.[1] But when, in pre-clinical trials, we extrapolate from such findings, via certain assumptions about dosage ratios, to the median lethal dose for humans, we are implicitly committing ourselves to a general hypothesis—about the drug's toxicity for mammals—which has only a certain measure of evidential support. For example, the hypothesis may be that with 95 per cent confidence limits the median lethal dose of a certain drug for any species of mammals is such-and-such a function of that species' mean weight. Here, in judging how well the hypothesis is supported, we can get no help from the mathematics of statistical theory. We have to gauge this support by reference to the variety of mammalian species used in the tests, the variety of methods used to administer the drug, the variety of other drugs administered in combination, and so on, and it is for so gauging evidential support that we need a confirmation-metric rather than a textbook of statistical method.

---

[1] Cf. *Experimental Chemotherapy*, ed. R. J. Schnitzler, F. Hawking, 1963, and *Quantitative Methods in Human Pharmacology and Therapeutics*, ed. D. R. Lawrence (1959).

No doubt it would be possible to bring any one or more of these factors within the scope of some statistical strategy. We could set out to determine, say, the *probability* that a drug tested safely on both cats and rats would be equally safe when administered to human beings. But any hypothesis about such a probability would in turn be exposed to assessment of its support from the evidence, e.g. in relation to the variety of drugs so tested, the variety of other drugs administered in combination, and so on. The need to employ a confirmation-metric for eliminative induction at some point does not begin to disappear until more or less the whole range of factors bearing on such experimental situations is well understood and the laws controlling them are well known—until, in other words, we are no longer dealing with a relatively primitive, but with a relatively advanced science.

There is an analogous situation in generative grammar. A hypothesis about the generative grammar of a particular language asserts, roughly, that any string of morphemes generated by such-and-such a set of rules is a grammatical sentence of the language. In the absence of some comprehensive fundamental theory, such as an adequate neurophysiological theory of the brain might conceivably be, a hypothesis of this kind has to be tested against native-speakers' intuitions of grammaticalness when the rules in question are applied over a variety of morpheme-classes within their scope; and testing on, say, a transitive and an intransitive verb may well provide more reliable results in a particular kind of case than testing on two different types of transitive verb. Here again, therefore, we have a case where comparative, and perhaps even quantitative estimates may need to be made of degrees of evidential support afforded by eliminative induction, and in this case the hypotheses concerned are specifically stated to have infinite domains. Correspondingly it is hardly surprising to be told[1] that assessments of grammaticalness are resistant to statistical support. For it is quite clear that generative grammar conforms to the conditions shown above to exclude any probabilistic assessment of evidential support. It conforms to the Instantial Comparability Condition because the relative weight of evidential support for an individual sentence's grammaticalness, in virtue of having such-or-such a structure, varies directly with that for the relevant rule or rules and vice versa: the grammaticalness of a string is to be judged only in the light of its supposed method of generation. Moreover the consequential conformity of generative grammar to the Instantial Conjunction Condition is illustrated by the way in which no-one supposes two applications of a set of generative rules to be any less supported by relevant evidence than a single application. And conformity, in ideal circumstances, to the ordering requirement included in the Evidential Relevance Condition is illustrated by the hierarchy of grammatical categories projected by Chomsky from a generative grammar, since this hierarchy may be regarded as presenting, in ascending order of subtlety, a series of tests for hypotheses about generative rules in that particular field.

---

[1] *Experimental Chemotherapy*, p. 30.

6

A few other possible objections remain. It is often claimed, for example, that rational credibility is best measured informally by betting odds and that since confirmation-functors should measure rational credibility their syntax must therefore be probabilistic. But betting odds are no use at all as a credibility-measure for many hypotheses of the kind with which I am here concerned. If a fair wager is to be made between A and B, it must at least be possible to describe both some circumstances in which the wager will be said to have been won by A and lost by B and also some circumstances in which it will be said to have been lost by A and won by B. Both sides must have some conceivable chance of winning. Suppose A bets on the falsehood of H and B on its truth. If H is a singular statement of observable fact it is easy enough to draw up the terms of the wager in such a way as to allow both A and B a conceivable chance of winning. But if H is a universal statement of observable fact with an infinite domain B has no conceivable chance of winning his bet. Betting odds are no use at all as a measure of rational credibility in such a case, and if we are to have any informal measure of justifiable partial belief here it must take some other form. For example it might be a system of preference-ratings between hypotheses, as in the example of the pre-clinical trials. Or it might be a scoring-system analogous to the marking or grading of students' performances in school or university examinations, where the student whose asserted beliefs are best supported by the available evidence scores highest. Assessment of a linguist's success in investigating the generative grammar of a language might be considered analogous in this respect to the assessment of a student's grammatical knowledge, and just as we sometimes have to revise a school marking-system when it fails to correlate with some other criterion of success, so too a confirmation-metric for grammar or any other science is equally subject to revision in the light of experience.

---

[1] N. Chomsky *Syntactic Structures*, 1957, pp. 16 f. I have discussed the application of confirmation theory to generative grammars at greater length with relevant references in 'On a Concept of Degree of Grammaticalness', *Logique et Analyse*, 30, June 1965, 141 ff. I also intend to discuss in another article the use of these confirmation-functors as a measure of content, information, meaningfulness or intelligibility.

More specifically, perhaps, it may be objected that sometimes the joint operation of two hypotheses is inhibited by factors that affect neither when operating alone. For example, taking the first of two particular drugs may alleviate one malady, taking the second may alleviate another, while taking them both together may be fatal. In such a case, it may be argued, the General Conjunction Condition cannot apply, because the conjunction of two hypotheses must not be supposed as well supported as either conjunct alone. But if in fact it is dangerous to take drug A and drug B together, then what we must suppose to be well supported by the evidence is not just that if A is taken a certain malady is alleviated, by that if A is taken *without B* a certain malady is alleviated, and similarly for B without A. The logical conjunction [illegible]ese two hypotheses may then safely be taken to be equally well supported. Of course, by the Consequence Condition (2), there is therefore at least as much support for the hypothesis that, if a man at the same time takes A without B and B without A both maladies are alleviated. But what state of health would it be false to assert of a man who achieved such a signally self-contradictory feat?

Perhaps it will be pointed out, by way of rejoinder, that whereas a man may be said to have some chance of beneficial effect from taking drug A, and some from drug B, he has none at all from taking both together. True, but my confirmation-metric is not concerned at all with chances or probabilities. It is concerned instead with assessment of evidential support from experimentally varied exposures to test. And this assessment is necessarily fallible since it is open to correction in the light of unfavourable experience. If the metric does permit the available evidence to give a high degree of support both to the hypothesis that A is beneficial and to the hypothesis that B is, and yet the conjunction of these two hypotheses later turns out to be false, then the metric is shown to be incorrect. It should have made even the previously available evidence give a high degree of support only to the hypotheses that A taken without B, and B without A, are beneficial. When we discover that barbiturates and alcohol go ill together we discover something not only about barbiturates and alcohol but also about the degree of support obtainable for certain sorts of hypotheses from certain sorts of test.

It might also be objected that, since I have accepted the Equivalence Condition (1), this kind of confirmation-theory is bound to be hit by Hempel's paradoxes.[1] But in fact it escapes them with particular ease. Certainly, if a hypothesis H scores *n*, say, according to a metric of this kind, on given evidence, then so too does any contrapositive equivalent of H. But there is no *a priori* need to suppose that according to such a metric the sort of evidence that generates Hempel's paradoxes should give any significant degree of support to a hypothesis. In terms of Hempel's example, we may suggest instead that experience has given our metric no warrant to assign any substantial degree of support to the hypothesis 'All ravens are black', or to its contrapositive, from the report that a non-black non-raven has been sighted, even though it has been sighted in circumstances in which the sighting of a black raven would lend substantial support to the hypothesis. It is most unlikely that ornithologists have any good reason to dispute this suggestion. Similarly, to revert to the pharmacological paradigm, we may suppose our hypothesis H to be something like the following: for all x and y, if x and y are two otherwise well-matched classes of mammals but the members of x are inoculated intravenously with the drug under test while the members of y are not, then x does not exhibit a significantly higher frequency of toxic symptoms than does y. If E states that a group of ten rats has been inoculated and then found not to differ significantly from a matched control group of ten uninoculated rats, the proposed confirmation-metric ensures that $c[H, E] = n$. But, if E′ states (analogously to the report of a non-black non-raven) that two groups of rats have been found to differ significantly from one another in frequency of toxic symptoms and that either both have been inoculated with the drug under test, or neither has, or they are not well-matched, our confirmation-metric would presumably be well justified in giving us some very small number, or even just 0, as the value of $c[H, E']$.

---

[1] C. G. Hempel, 'Studies in the Logic of Confirmation', *Mind*, 54, 1945, 1 ff.

In short, Hempel's paradoxes arise in connection with Nicod's criterion for a purely qualitative concept of confirmation. They may be regarded as helping to teach us the inadequacy of such a concept and its commonly associated conditions and criteria. But a quantitative or comparative concept of evidential support need not be hit by them at all, so long as it is associated with an empirically based metric—a metric that is sensitive to our actual experience of the value of different kinds of evidence.

Finally, I wish to emphasise again that I am concerned only with the logic of evidential support, not with its various metrics (which I conceive to be problems for scientists themselves, not for philosophers), and only with estimates of evidential support obtainable via the kind of eliminative induction that I have described. For example, perhaps two different functions may be proposed as a solution to the problem of how a certain two physical parameters are related, both functions may fit equally closely all the available values of these parameters, both functions may be equally simple, and yet both proposals cannot be correct because they lead to incompatible predictions for values of the two parameters which are at present not experimentally obtainable. Clearly, the two rival hypotheses here do not conform to my General Conjunction Condition (10) if they are taken to derive some positive degree of support from the available evidence. But then they do not fit the Negation Condition (11) either. Nor is this surprising, since the evidential support they enjoyed is a mode of enumerative rather than eliminative induction[1] a mode of deriving support from some instances for all instances rather than from some kinds of instance for all kinds of instance. The amount of support each hypothesis is supposed to have is apparently assessed solely by reference to the fact that the evidence fits this hypothesis, and quite without reference to the fact that it also fits a rival hypothesis equally well. Enumerative induction employs only the former method of assessment, because it does not aim, by deliberate variation of circumstance, to choose between competing hypotheses.

Indeed, in view of the intensely controversial nature of much that has been written in this field we might well expect there to be more than one logic of evidential support. But until all the relevant concepts have been thoroughly examined and systematically articulated we shall scarcely be in a position to decide their several rôles and relative importance.

(*to be concluded*)

---

[1] It would be part of a larger undertaking than my present one to show, as I suspect can be shown, that enumerative induction is just as much tied to a probabilistic conception of evidential support as eliminative induction is resistant to it. (In general, probabilistic theories of eliminative induction tend to invoke speculative principles of limited independent variety and fail to represent the essentially corrigible and empirically testable character of assessments of evidential support by eliminative induction.)

COULD A MACHINE PERCEIVE?*

ALAN GAULD

I

THE belief that it is possible to construct a man-like machine is by no means new. Daedalus of Crete was alleged to have built moving statues so agile that it was necessary to tie them down in order to prevent them from running away, and Pope Sylvester II, Albertus Magnus and Roger Bacon were widely thought to have made marvellous talking heads from plans supplied by the devil:

> I tell thee, *Bacon*, *Oxford* makes report,
> Nay, *England* and the court of Henrie saies
> Thart making of a brazen head by art,
> Which shall vnfold strange doubts and aphorismes,
> And read a lecture in Philosophie. . . .[1]

The artificers who flourished after the Reformation were less successful in devising machines to simulate men—perhaps they did not care for the risks which their predecessors had run. Their most exalted achievement was probably the humble (but extremely realistic) automaton duck built by Vaucanson in the eighteenth century. It is only within the last twenty years or so that advances in electronic gadgetry have made the construction of a man-like machine once again seem a practical possibility. We know now that if we can describe any part of any individual's life-history, or even the whole of it, in a finite number of words, we can in principle programme a digital computer to parallel that life-history to any required degree of exactness.

Modern psychology is the junk-shop of the sciences; the most attractive goods in its window came originally from more prosperous establishments. In recent decades it has frequently cast eyes on the electronic gadgets mentioned in the last paragraph; and not a few psychologists have come to feel that a promising way of approaching the study of human behaviour may be to seek analogies between the

---

* Received 17.viii.65

[1] R. Greene. *The Honourable Historie of Frier Bacon and Frier Bongay*, Act 1, Scene ii.

workings of the brain and the operations of highly complex machines.[1] Critics of this viewpoint have commonly suggested that since there are certain highly abstract tasks (notably proving various mathematical theorems) which men can perform but machines cannot, a mechanistic approach to the understanding of behaviour must ultimately break down.[2] In this article I shall suggest that anyone wanting to design a man-like machine would encounter notable difficulties not merely in arranging that his machine should generate certain abstruse theorems, but even in ensuring its ability to carry out behaviour which would present no problem to the most ordinary citizen. Should these difficulties turn out to be insuperable it would follow that not a little of the work currently being done in the name of a scientific psychology is radically misconceived.

2

Some simple illustrations of the difficulties I have in mind can be obtained from the phenomenon which psychologists have variously labelled ' object perception ', ' the perception of things ', ' perceptual learning ', ' object learning ', ' cognition ', or even ' the perception of meaning '. The problem of object perception is sometimes expressed by asking how we come to be able to recognise a chair *as* a chair, a tomato *as* a tomato, and so forth. The traditional solution of this problem was that a sensation (say the round, red sensation produced in one's visual field by a tomato) will, if it occurs sufficiently often in a fairly constant context, come to arouse ' associations ' (with the taste and feel of a tomato, etc.) which form a background of expectations about how the object which produced the sensation is likely to behave. ' Perception ', said William James, ' differs from sensation by the consciousness of farther facts associated with the object of the sensation '.[3]

---

[1] A general idea of this kind of approach may be obtained from, e.g. E. A. Feigenbaum and J. Feldman (eds.) *Computers and Thought*, New York, McGraw-Hill, 1963; M. A. Arbib, *Brains, Machines and Mathematics*, New York, McGraw-Hill, 1964; H. Borko (ed.) *Computer Applications in the Behavioral Sciences*, London, Prentice-Hall International 1962; B. F. Green, *Digital Computers in Research: An Introduction for Behavioral and Social Scientists*, New York, McGraw-Hill, 1963.

[2] See e.g. E. Nagel and J. R. Newman, *Gödel's Proof*, London, Routledge & Kegan Paul, 1959, pp. 100–1; J. R. Lucas, 'Minds, Machines and Gödel', *Philosophy*, **36**, 1961, 112–27; and for an attempt to counter the arguments J. J. C. Smart, *Philosophy and Scientific Realism*, London, Routledge & Kegan Paul, 1963, pp. 116–120.

[3] W. James, *The Principles of Psychology*, London, Macmillan, 1891, **2**, 77.

Nowadays this ' enrichment ' theory of perception is often translated into the fashionable jargon of behaviour theory or of neuropsychology; ' descriptively ', says L. Postman, ' perceptual learning *is* the attachment of new responses, or a change in the frequency of response, to particular configuration of stimuli '.[1] Since ' instrumental conditioning machines '—machines which will ' learn ' to attach new responses to particular configurations of stimuli—can readily be built, one might very well suppose that the construction of a machine which could ' perceive ' in the sense in question would present no difficulties of principle.

The obvious objection to such a standpoint is that perceiving is a special inward activity which no machine could possible duplicate. I shall, however, neglect this objection for the moment, and shall suggest instead that in order to perceive an object *as* an object of a certain sort one has in many instances to exercise concepts of a very advanced level; so advanced, in fact, that it is hard indeed to suppose that anyone could even in principle design a machine which might properly be said to 'possess ' them.

3

What, for instance, would someone have to do in order to convince us that he was capable of (say) perceiving a pound note as a pound note? He would certainly have to do more than just make consistent responses to certain stimulus configurations. No one would dream of saying that a rat who consistently ate pound notes whenever he met with them was seeing them as pound notes. And it is quite obvious why. To see a pound note *as* a pound note is to see it as something of value, as something one can exchange, as legal tender: and value, exchange and tender are high-level notions which a rat cannot be supposed to possess. Similarly, even though a certain human being always said ' why there's a pound note ' correctly in the presence of pound notes (and never said it on other occasions), we might in certain circumstances still be reluctant to say that he had learned to perceive pound notes *as* pound notes. If his other behaviour showed clearly that he had no idea that pound notes can be exchanged for other forms of money or for goods, or that they can be paid into bank accounts or used for bribes or given as wages we should probably say that he simply saw pound notes as pretty patterns, and not as pound notes, as bank notes of a certain value. In taking someone's exclamation of ' why there's a pound note ' as an indication that he sees a pound note *as* a pound note, we have *necessarily* to assume that he possesses these rather high-level concepts; his possessing such concepts is intrinsic to his being able to see pound notes *as* pound notes.

---

[1] L. Postman, 'Association Theory and Perceptual Learning', *Psychological Review*, 1955, **62**, 441.

Of course there is not in fact a sharp dividing line between, on the one hand, animals consistently making certain responses to particular configurations of stimuli, and, on the other hand, human beings in full possession of high-order concepts like money, payment or exchange, and hence capable of developed object perceptions. It is clearly possible to think of all sorts of intermediate cases. A chimpanzee who will learn to work for poker chips which he can insert in a food vending device[1] might be said to be a step nearer possession of monetary and related concepts than a rat who merely makes a consistent discriminatory response to pound notes; but we would hardly say that the chimpanzee actually possessed these concepts unless he showed signs of attaching *generalised* value to the poker chips. Contrariwise, children whose monetary and related concepts are imperfectly developed—who, for instance, tend to confuse the value of an object with its physical size[2]—might none the less have sufficient grasp of such concepts for us to admit their ability to see pound notes as pound notes. The point is that only when someone begins to develop high-level monetary and related concepts can we say that he is beginning to be capable of seeing pound notes as pound notes.

Now the high-level concepts in question seem to be ones which could not be acquired by simply ' abstracting ' common properties from the various physical situations which are said to be instances of the concept[3]; as (it might be supposed) the concept of cat can be arrived at by abstracting from certain physical events those features, qualitative or relational, which all cats have in common. Do all instances of (say) ' exchange ' share a common physical feature, either qualitative or relational? Hardly; think of primitive people exchanging goods for beads or wives for cattle; of someone signing a cheque for shares; of services rendered in exchange for lands; of school children swapping toys and so forth. The concept of exchange would seem susceptible of definition only in terms of other high-level concepts—e.g. promise, bond, agreement, commodity, transfer, benefit, value. Possession of concepts like ' exchange ' certainly cannot consist just in making certain responses to all situations sharing a given physical feature; and hence *a fortiori* the ability to perceive pound notes as pound notes cannot *just* consist in a tendency to make certain responses to certain stimulus configurations. We may, therefore, at any rate claim that, whether or not it is possible to design a machine which could carry out the activity called ' object perception ', the traditional analysis of that activity is, at least in some cases, very much wide of the mark.

---

[1] J. T. Cowles, ' Food Tokens as Incentives for Learning by Chimpanzees ', *Comp. Psychol. Monogr.*, 1937, 14, no. 7.

[2] J. S. Bruner and C. C. Goodman ' Value and Need as Organizing Factors in Perception ', *J. Abnormal Soc. Psychol.*, 1947, 42, 33-44.

[3] On the difficulties of ' abstractionism ' cf. Peter Geach, *Mental Acts*, London, Routledge and Kegan Paul, n.d., pp. 18-44.

4

It is easy to think of parallel examples, and there is no point in spending much time on them. At what stage in a child's development would we agree that he had (say) become capable of seeing a policeman *as* a policeman, a bookmaker *as* a bookmaker, a priest *as* a priest, a playing-card *as* a playing-card, a postage stamp *as* a postage stamp, a diamond *as* a diamond, a Union Jack *as* a Union Jack, a Victoria Cross *as* a Victoria Cross? Not before we could discover in his general behaviour some mastery of such concepts as war, enemy, patriotism, religious teaching, sport, sportsmanship, game, nationality, message, law, communication, bet, odds, monetary value, bravery and courage. The possession of the concepts is *essential* to the perceptual abilities; a policeman *is* someone who enforces the law; a priest *is* someone who delivers religious teachings; a playing-card *is* something which is used in games. If a person possessed, for instance, no legal or religious concepts, he could perceive policemen and priests only as people who wore funny clothes and frequented special places.

It might of course be suggested that the instances of object perception to which the possession of such concepts is essential are really rather rare. One could perceive a chair as a chair, a cat as a cat, a tomato as a tomato, or a thousand other commonplace objects as what they are, without having a more advanced conceptual equipment than a rat might be supposed to possess. From a behavioural point of view, to come to see a tomato as a tomato (for instance) just is, as Postman put it, to exhibit the attachment of certain responses (grasping, slicing, eating, etc.) to certain configurations of stimuli; and such behaviour is in principle no different from the behaviour of a rat who learns to 'recognise' peanuts, who learns, that is, to approach certain sources of stimulation and nibble.

These suggestions could be answered in two ways. First of all one might point out that the examples of object perception which I have taken are by no means recondite—what could be less unusual than seeing a diamond, a pound note, a Union Jack, a playing-card or a policeman? But secondly, one might add that these examples were chosen simply for convenience of exposition; exactly the same principles of analysis might perhaps be applied even to some among the simplest cases; say to the case of perceiving a tomato as a tomato. After all, it is plainly nonsense to assert that learning to perceive tomatoes as tomatoes simply consists in coming to approach tomatoes and eat them; this would not distinguish learning to perceive a tomato as a tomato from learning to perceive it just as edible, or from coming to misperceive it as a sort of meat (like a kidney perhaps). Nor would someone's learning to say 'tomato' whenever faced with a tomato be a complete guarantee that he has learned to perceive tomatoes as tomatoes. If his other behaviour gave no hint that he knew tomatoes are edible, or that he knew they are vegetable rather than animal or mineral, or that he knew they are or had been alive, we should have to deny that he could see tomatoes as tomatoes. But it is surely very doubtful whether even the ordinary man's concept of 'being alive' can be analysed in terms of the acquisition of certain responses to certain configurations of stimuli.

The crucial question now arises, can we in principle design a machine which could properly be said to 'possess' such high-level concepts? Upon the answer to this question, of course, depends the answer to the question of whether or not we can in principle design a machine which could properly be said to carry out the activity of object perception.

5

The first problem which we must tackle is clearly how to define the word 'machine'. To give a general definition of a machine would undoubtedly be very difficult; but there does seem to be widespread agreement that we can define an important sub-class of machines, that of 'finite-state' machines, in the following general way.[1] A 'finite-state' system is a system which must at any given moment be in one out of a finite number of possible states. Now if such a system is to be a *machine*, an interconnected system rather than a pointless jumble, there must be rules which specify how its various possible states are to succeed each other. And since in many machines the order of succession of possible states may be influenced by 'inputs', that is, by changes of state more or less arbitrarily imposed from outside the system, the rules in question will often have to include rules specifying the effects of the various possible inputs. We could set down all the rules for any given machine in the form of a 'machine-table', which lists for each possible state of the machine what the machine will do next, and especially what it will do next if any given one of the possible inputs should arrive. Any physical system whose operations can be described by such a machine-table constitutes a finite-state machine.

This general way of defining a machine can in principle be extended to cover more complex cases—for instance, machines whose states succeed each other in accordance with probabilistic rules,[2] or machines in which changes of state themselves produce inputs—provided always that the systems concerned have a finite number of discrete possible states.

Any machine for which a machine-table can be written, that is to say any finite-state machine, can be embodied in a digital computer, which is a 'universal' finite-state machine; and since it is possible

---

[1] It should be noted that under this definition natural as well as artificial systems can qualify as 'machines'. This sort of definition of a machine (here given very informally) appears in many variants. Most of them are probably derived ultimately from A. M. Turing's celebrated paper 'On Computable Numbers, with an Application to the Entscheidungsproblem', *Proc. London Math Soc.*, 2nd series, 1937, 42, 230-265. Informal versions are given, e.g. by Turing, 'Computing Machinery and Intelligence', *Mind*, n.s. 1950, 59, 439-42; by N. Moray, *Cybernetics: Machines with Intelligence*, London, Burns and Oates, 1963, p. 107; by W. Ross Ashby, 'Principles of Self-Organizing Systems', pp. 261-2 in Foerster and Zopf (eds.), *Principles of Self-Organization*, Oxford, Pergamon, 1962; by W. Ross Ashby, 'Simulation of the Brain', pp. 458-9, in Borko (ed.), op. cit. Somewhat more formal versions can be found in M. A. Arbib, op. cit. pp. 7-10, and in W. Ross Ashby's *An Introduction to Cybernetics*, London, Chapman and Hall, 1956, pp. 9-24 and 42-48. A useful survey is R. McNaughton 'The Theory of Automata,' in F. L. Alt (ed.), *Advances in Computers*, New York, Academic Press, 1961, 2, 379-421. A full mathematical statement (with references) is given by A. Gill, *Introduction to the Theory of Finite-State Machines*, New York, McGraw-Hill, 1962, pp. 1-13.

[2] Cf. what Ashby, op. cit. pp. 225-35, says about 'Markovian machines'.

(though sometimes implausible) to *regard* any machine as a finite-state machine,[1] one can on a digital computer in principle formally parallel to any required degree of exactness the behaviour of any machine whatsoever.

6

It would be perfectly easy to formulate the machine-table for (let us say) a miniature Jones-machine. Suppose that the Jones-machine has four possible states—resting, drinking, winking and ordering beer; and eight possible inputs—bar *plus* barmaid, beer, or landlord, either one by one or in various combinations. Then for each possible combination of input and machine-state we can tabulate what state the machine will move to next:

	Machine-States			
Inputs	M1 *Resting*	M2 *Drinking*	M3 *Winking*	M4 *Ordering Beer*
I1. Bar	(no change)	M1	M1	M1
I2. Bar+Beer	M2	(no change)	M2	M2
I3. Bar+Barmaid	M4	M3	(no change)	M4
I4. Bar+Landlord	M4	M1	M1	(no change)
I5. Bar+Beer+Barmaid	M2	(no change)	M2	M2
I6. Bar+Barmaid+Landlord	M4	M1	(no change)	M4
I7. Bar+Beer+Landlord	M2	(no change)	M2	M2
I8. Bar+Beer+Landlord+Barmaid	M2	(no change)	M2	M2

The machine specified by this table would simulate some parts of the real Jones's behaviour quite closely. When it had neither beer, barmaid or landlord, but only the empty bar, in front of it, it would rest. If given beer it would drink it whether or not anyone else were present. If the barmaid were there when it had finished its beer it would begin to wink at her; but if the landlord came in it would stop winking, pause for a moment, and order more beer. And so on. Some people might well be tempted to say that such a machine 'possessed' in rudimentary form some of the concepts which Jones himself possesses, and especially, of course, the concept of beer.

---

[1] Cf. Ashby, op. cit. pp. 9-10, 28.

7

Could we go further, and draw up a machine-table the embodiment of which in hardware would constitute a machine that might properly be said to possess high-level concepts like game, sportsmanship, exchange, law, and so on? It is obvious that the rules of such a machine would have to be far more complex than the rules of the miniature Jones-Machine just described; but it is not at once obvious that they could not be set down in the approved form. For instance, to take an example made famous by Wittgenstein, if Jones possessed the concept of game his behaviour in that respect might be exhaustively described by a list of such rules as I(nput) a→S(tate)x, Ib→Sy, and so on, where the inputs are games or slices of games and the states represent vocalisations such as ' well played, chaps ', or ' jolly good show '. Of course a very large number of very complicated rules might be required to cover (say) Jones's concept of sportsmanship. Suppose for instance that he received Input Ic (fielder with ball in hand, umpire saying ' not out ', yet batsman leaving the wicket) and was in State Sn (sitting, mouth empty), the rule might be Ic+Sn→So (where So is crying ' a sporting gesture, sir '); but if he received Input Ic whilst in state Sp (sitting, eating a sandwich), the rule might be Ic+Sp→Sq (clapping the hands). If we allowed an indefinite, but not infinite, number of additions to the rules, and to the complexity of the rules, we could eventually reach a set of rules which would more or less completely describe Jones's behaviour in so far as it exhibited a concept of sportsmanship—at least if we permitted a touch of randomness here and there. Jones may not always have had this set of rules, and he may perhaps change his rules in the future; but nonetheless these rules describe how he acted in sporting situations during a definite segment of his immediate past. Since we can put these rules into a machine-table, we can, at any rate in theory, build a machine which would undoubtedly ' possess ' Jones's concept of sportsmanship.

8

So, at least, a radical machine theorist might be prone to allege. But even if we had drawn up a machine-table which fully described Jones's behaviour in sporting situations over a large part of his life, would the embodiment of this table in a digital computer really give us a machine which could be said to ' possess ' Jones's concepts of game and sportsmanship? It would be very, very odd to make such a claim. Any person of moderate sophistication who looked down the list of rules in the machine-table, and who understood to what external situations and to what bodily positions the various headings in the input and machine-state lists corresponded, would at once be able to see that the rules concerned belonged together, that they possessed a certain unity. Yet in what could this unity consist? Certainly not in physical similarities among the external situations and among the bodily states to which the rules refer. Have we simply learned to say of each of the actions-in-a-context prescribed by the rules ' this is an exercise of the concept of game ' or ' this is an exercise of the concept of sportsmanship?' How many people teach such rigmaroles to their children, or can remember learning them from their parents? And if they did learn them from their parents, how did their parents come to recognise the unity among all these disparate physical situations? Surely there is no escaping the conclusion that so far from these sets of rules constituting our concepts of game and of sportsmanship it is only because we already possess concepts of game and of sportsmanship not reducible to such sets of rules that we can recognise the physical situations covered by the (supposed) rules as instances of the exercise of the concepts of game and of sportsmanship.

9

There is another line of argument which suggests that possession of sporting concepts cannot be analysed into a mere propensity to respond in accordance with a set of rules relating possible inputs to possible changes of bodily state. Suppose that we have observed Jones's behaviour continuously since the day he was born, and that we have been able to produce a complete list of the principles of behaviour which are supposed to make up his concepts of game and of sportsmanship; and suppose further that we have embodied these principles in a miniature Jones-machine. We then present both Jones and the Jones-machine with a series of novel game-situations; novel game-situations being by definition ones which bear no special physical resemblances to the game-situations Jones and the Jones-machine have encountered in the past. The Jones-machine will obviously not be able to recognise the novel games as games, because the necessary input-change of state rules will not have been built into it. But what if Jones himself has no difficulty in recognising the novel games as games, and in singling out from within them deeds of outstanding sportsmanship? There would seem to be only three possible kinds of explanation, viz. (1) that through some hereditary memory Jones already has the requisite additional rules ' built in ' to his nervous system; (2) that he has elicited the rules telepathically from the psychologists experimenting upon him; and (3) that he is acting in terms of concepts of game and of sportsmanship not reducible to sets of rules which could be built in to a machine. A theory about hereditary memory would these days hardly be entertained for a moment by anyone; and in any case it would involve the ludicrous assumption that the ' novel ' games concerned were known to Jones's ancestors. The second possibility would at best be hard to reconcile with machine theory. This leaves us only the third possibility. Accepting this view would force us to admit that sporting concepts cannot be acquired by learning (as a machine might do) a set of rules about how to react in each of a given series of physical situations. The concepts may in some sense be created out of such rules; but they also transcend and unify them.

Whether or not people can recognise games and acts of sportsmanship physically unlike any that they have seen before is a question which could in theory be answered by observation and experiment. In practice, of course, the issue cannot be decided in this way; we cannot restrict the lives and activities of our fellow human beings sufficiently to be able to devise a crucial test. Nonetheless there seem to me to be considerations which rather strongly indicate that we can without special training extend our concept of game (and *a fortiori* that of sportsmanship) to cover any number of instances physically unlike those which we have previously encountered. For example not very long ago I watched two small children indulging in the following pastime. Across a concrete garden path about a yard in width they laid blades of grass end to end; complete bridging of the path (which was difficult to achieve owing to constant gusts of wind) seemed to constitute a ' success ' and led to a short pause before activities were resumed. Now I had no hesitation in assuming that the children were playing a game, and this despite the fact that I certainly cannot recollect seeing any game in the least bit similar played anywhere before. Furthermore there is no doubt that the children themselves would have referred to their goings-on as a 'game ', although, so far as I am aware, they had devised it for the first time that afternoon, doubtless because of the state of the wind and the fact that there were grass cuttings around. Now if children can devise games physically unlike any game that they or other people have seen before (and children, like the organisers of *Beat the Clock*, are constantly inventing the queerest games), and if those children, and other children, and adults, can recognise the games as games, then of course it follows, from the argument of the last paragraph, that the children and adults concerned must possess concepts of game (and possibly of sportsmanship too) which cannot be reduced to mere sets of rules about how to behave in each of a number of physically definable situations.

10

Indeed that possession of the concept of game cannot be reduced to such a set of rules seems to follow just from the definition of the word ' game '. Consider the plight of a social anthropologist studying some outlandish tribe or bizarre sub-culture, and puzzled as to whether or not a certain piece of behaviour is a game in the occidental sense. Surely the proper way for him to decide this question would not be to rack his memory in an effort to recollect whether he or anyone else had ever seen anything physically like the activity concerned before; but to inquire whether or not it had rules and a criterion of success, was pursued for diversion, required skill or luck, and so on. A game simply *is* an activity which is pursued for diversion, has rules and a criterion of success, etc. (even the grass-blade game described above had rules—such as that you must use grass-blades and not Michaelmas daisies). Now any attempt to define notions like ' rule ', ' criterion ', ' success ', ' diversion ', ' skill ' and ' luck ' in terms of rules relating organismic states to physical inputs would present us afresh with exactly the sort of difficulties we have already encountered. It seems as though the concept of game must systematically elude any definition of a kind which could be embodied in a machine-table.

11

People have of course built ' games-playing ' machines—machines whose operations formally parallel the moves in, for instance, games of chess or of tic-tac-toe. But such machines, however versatile, no more possess the concept of game than a cash-register which always gives the right change possesses the concept of honesty. Someone who possesses the concept of game must at least be able to recognise a wide variety of games for what they are; just as someone who has the concept of honesty must at least be able to pick out honest actions from dishonest ones in all sorts of different contexts.

12

It would be very satsifying if, from our demonstration (or would-be demonstration) that to possess sporting concepts is not simply to follow rules of a kind which could be embodied in a machine, we could pass directly to the claim that machines cannot possess such concepts, and hence, of course, to the claim that a machine could not perceive a playing-card *as* a playing-card or a checker-board *as* a checker-board. Unfortunately to show that possession of sporting concepts does not consist in a tendency to follow sets of rules which could be embodied in a machine hardly gives us grounds for asserting that a machine could not possess these concepts. To drive this further assertion home we should need to make clear what it actually *is* to possess such concepts, and then to give convincing reasons why a machine could not properly be said to possess them. The task would be considerable, and can hardly be undertaken in the space that remains.

Nonetheless I think that we can at this juncture justly address the following remarks to any sanguine machine theorist. You have adopted a theoretical framework which seems *prima facie* to force upon you a kind of analysis of what possession of a concept consists in which we have shown to be, in the case of certain concepts at least, quite untenable. You must, it appears, analyse possession of a concept in terms of rules relating a machine-organism's possible inputs to the various states which it is capable of assuming.[1] This analysis is misconceived in the case of sporting concepts, and could be shown to be misconceived in the cases of numerous other high-level concepts too. The failure of such concepts to fit within your chosen framework completely vitiates your claims to be able to give useful explanations of human behaviour; for if a machine cannot possess these high-level concepts it cannot properly duplicate the commonplace and universal human activity of object perception. If we are to continue to take seriously your large claims that yours is the royal road to the understanding of human behaviour, we must require you to produce a fresh analysis of what it is to possess such concepts—an analysis *not* couched in terms of rules of the kind that can be set down in a machine-table—and then to convince us that, in specified circumstances, a machine could *without self-contradiction* be said to possess them. If you fail in this task, or decline to undertake it, we must positively refuse to waste further time in studying your speculations. We have enough to occupy us as it is; for life is short, and the complexities of human nature are inexhaustible.

---

[1] Those machine theorists who have written about what it is to possess a concept do indeed generally adopt an analysis which fits into this pattern; so do 'behaviour' theorists, who often share machine theorist's presuppositions. A couple of examples will perhaps suffice. In a celebrated paper on 'Programming a Model of Human Concept Formation' (reprinted in Feigenbaum and Feldman, op. cit. pp. 310-325), E. B. Hunt and C. I. Hovland say: 'Church . . . has offered a definition [of what a concept is] which has been accepted, implicitly, by psychologists who perform " concept learning " experiments. Church's argument is that any given symbol (or *name*) can be attached to the members of a set of objects. For any arbitrary object there exists a rule concerning the description of the object, a rule which can be used to decide whether or not the object is a member of the set of objects to which the name applies. The decision rule is the concept of the name, the set of objects is the denotation of the name.' C. E. Osgood, approaching the question from the point of view of a behaviour theorist, says 'Just what, in everyday language, is a concept? It is a common response (usually verbal) made to a class of phenomena the members of which display certain common characteristics' (*Method and Theory in Experimental Psychology*, New York, Oxford University Press, 1953, pp. 666).

13

I have tried to show that it is distinctly doubtful whether one could even in principle design a machine whose behaviour would force us to admit that (like a human being) it could perceive objects as objects of certain sorts; and I believe that similar arguments could be used to suggest that it is likewise distinctly doubtful whether one could design a machine which could truly be said to carry out the human activities of hoping, wishing, feeling, thinking, remembering, sympathising, intending or using language. Every normal person possesses concepts which are not susceptible of a reductive analysis; possession of such concepts is the chief asset which distinguishes a human being among his peers from an idiot in an asylum. If one wants to understand the behaviour of an intelligent human being one must first of all discover the concepts in terms of which he is acting; if necessary one must acquire, or re-create, the concepts for oneself. No other approach can succeed; for it is *these* concepts which (in some sense) regulate his behaviour. They cannot be analysed reductively or otherwise dissolved; they can only be explored and elucidated. Indeed the elucidation of concepts must generally precede empirical inquiries if the latter are not to run the risk of proving vacuous. Suppose, for instance, that someone propounded the thesis that the strength of the positive sentiment people feel for each other increases with the frequency of their social interactions.[1] At first sight this looks like a simple empirical proposition which could be tested by observation and experiment. But elucidation of the concepts of sentiment and of social interaction held by individual members of our society might show that part of what people mean by saying that they have a strong positive sentiment for someone is that they often have social interactions with him. The proposition that sentiment varies with interaction, which seems empirically testable, might in the end turn out to be true by definition.

14

Too many of today's psychologists take it for granted that a complete mechanical explanation of human behaviour could in principle be achieved. Some of these writers are machine theorists pure and simple —people who think that engineering principles can be directly applied to the explanation of behaviour; others would call themselves 'behaviour theorists' and would prefer a neo-Pavlovian terminology. But the upshot is the same. Of course no-one would dispute that many problems in animal behaviour, and some in human behaviour, can profitably be approached in this way; but these problems are for the most part of slight practical consequence. How unfortunate it will be if all psychologists wholeheartedly embrace a view of their subject-matter which will eventually force them, if they wish for any success at all, to investigate reaction times and rats instead of the social life of aspiring humanity; to do odd jobs for physiologists instead of devloping and expounding their own proper study; to frequent, in short, multitudinous back alleys of no moment in preference to the broad highway where all men pass.[2]

Department of Psychology
University of Nottingham

---

[1] A thesis somewhat resembling this figures prominently in G. C. Homans's *The Human Group*, London, Routledge and Kegan Paul, 1951.

[2] The author wishes to thank Dr. R. B. Joynson and Mr. A. J. Watson for their helpful criticisms of a draft of this paper.

NOTES AND COMMENTS

A Paradox of Information[*]

For the sake of simplicity I shall confine myself initially to two interpretations of the probability calculus: the *logical* interpretation and the *statistical* interpretation.

Let A and B be statements of some suitable language. Then I write

$$P(A, B) = r$$

for ' the (logical) probability of the statement A, given the information statement B, equals r '.

Let a be some possible experimental result on a chance set-up: and let A state that the event a occurs. Then I write

$$p(a) = r$$

for ' the (statistical) probability of a, on the given chance set-up, equals r '.

We shall consider here the case in which a is the falling 5-uppermost of this die on its next throw. We will assume that the die is fair; that is, we will make the following *factual assumption* (α):

$$p(a) = 1/6 \tag{α}$$

Then A will be the statement that says that a occurs: in our case, the statement ' In the next throw this die will fall 5-uppermost '.

We now assume that ' $p(a) = r$ ' may be used as an *information* statement; that is, that among the statements $B, \ldots$ there will be, for every r ($0 \leq r \leq 1$), statements

$$p(a) = r$$

which we denote by ' E_r^a ' (' E ' for *evidence* or *information*).

Under this assumption the following principle seems intuitively necessary:

$$P(A, E_r^a) = r \tag{1}$$

By way of illustration we will show that (1) yields a further highly plausible principle. We write ' $P(A)$ ' for the probability of A on *tautological* information (the absolute logical probability of A). In particular we consider the following tautological information

$$p(a) = p(a)$$

which we denote by ' $E_{p(a)}^a$ '.

---

[*] This note was distributed to participants in the *International Colloquium on the Philosophy of Science* held in London, 11th-17th July 1965.

At once, by substitution in (1) we obtain

$$P(A) = P(A, E^a_{p(a)}) = p(a)$$

which in view of the fact that A asserts that a occurs is (at least *prima facie*) very reasonable.

We will now show that the apparently necessary principle (1), together with two trivial and unexceptionable formulae of the probability calculus, yields a contradiction.

For we have by the calculus,

(2) $$p(a) = 1/2 \equiv p(a) = p(-a)$$

(where '$-a$' denotes the event that occurs if and only if a does not occur). and therefore,

(3) $$E^a_{1/2} \leftrightarrow E^a_{p(-a)}.$$

Moreover, we have by the probability calculus, which permits substitution of logically equivalent statements for one another,

(4) $$P(A, E^a_{1/2}) = P(A, E^a_{p(-a)}).$$

This may still sound reasonable: yet we have, by substitution in (1),

(5) $$P(A, E^a_{1/2}) = 1/2$$

and

(6) $$P(A, E^a_{p(-a)}) = p(-a),$$

and therefore, combining (4), (5) and (6),

(7) $$p(-a) = 1/2,$$

whence, by (2),

(8) $$p(a) = 1/2.$$

We now obtain from (8) and the *factual assumption* (α),

(9) $$1/2 = 1/6.$$

The result can immediately be generalised by interpreting '$p(a) = r$' not as a statistical statement but as any other sort of probability statement. It may, for example, be interpreted as a statement of so-far observed frequency, or even as a logical probability statement, with a as a *translation* of A into some suitable language.

Under the former re-interpretation, (1) becomes one form of the so-called '*straight rule*' (more precisely, Reichenbach's rule of induction). Since our argument remains fully applicable, the straight rule, at least in this form, has been shewn inconsistent.

There are a number of further generalisations: for example, we may replace the 'r' on the right-hand side of (1) by certain numerical functions, '$g(r)$', of r, without essentially disturbing our argument. Only such generalisations would be of interest that lead to an interpretation of (1) which has been proposed by some worker in the field as intuitively satisfactory, or which is in itself intuitively attractive.

In the form in which the information is interpreted statistically, the present result may be said to support a suspicion repeatedly voiced by Professor Popper.[1] I should like to thank him for the great interest he has shown in discussing the problem of this note with me.

London University DAVID MILLER

A Comment on Miller's New Paradox of Information

The important and indeed brilliant discovery described by Miller in the preceding note will be discussed here with a view to (I) explaining some of its main points, (II) and (III) expounding it formally, and (IV) applying it to Carnap's probabilistic theory of induction.[2]

I

Let 'E^a_r' be the name of some (evidence-) statement like

'The observed (relative) frequency of the event a equals r'

or

'The statistical probability of the event a equals r'

or in symbols, '$fr(a) = r$' or perhaps '$p(a) = r$'

where a may be, say, the event of a throw with a certain die resulting in the number five turning up.

Let 'A' be the name of the statement (hypothesis) that the event a will occur again; for example, A may be the statement:

'The next throw with this die will result in the number 5 turning up.'

Let '$P(A)$' (or '$P(A, E^a_r)$') be some shorthand notation for

'The probability, or the degree of confirmation, of A'
(or '. . . of A in the light of the evidence E^a_r').

---

[1] 'Probability Magic, or Knowledge out of Ignorance', *Dialectica*, 11 (1957), pp. 354 ff., especially the end of section 1. See also 'The Mysteries of Udolpho: A Reply to Professors Jeffrey and Bar-Hillel', forthcoming in *Mind*, particularly section 10.

[2] Sections I and II of this note have been added to the note originally distributed together with Miller's note (and some other notes) to the participants in the *International Colloquium on the Philosophy of Science* held in London, 11th–17th July 1965. Apart from the insertion of section-numbers 'III' and 'IV', the two other sections have been left practically unaltered, except for an abbreviation of the derivation in the section now called 'III', and a brief comment on the avoidability of certain substitutions at the end of this section.

Then the following rules (Rc) and (Rcr) may seem to some people intuitively convincing, or at least very plausible:

(Rc) If E_r^a is true, then $P(A) = r$

or perhaps,

(Rcr) If E_r^a is true, then $P(A, E_r^a) = r$.

The question whether these rules are generally acceptable will not be discussed here; indeed, I will assume, for the present purpose of discussing Miller's paradox, that these rules are valid rules of induction (or valid forms of what is now usually called *the straight rule* of induction).[1]

What Miller's paradox shows is the following:

If we try to write (Rc) or (Rcr) in a certain form which looks similar to (Rc) and (Rcr) but which makes *essential use of the probability calculus*, then we arrive at a contradiction, merely by using the rule and certain theorems of the calculus.

Let us first see what we cannot derive from (Rc) or (Rcr).

Assume that we are given that $E_r^a = E_s^a$ are inter-deducible; that is to say, we are given the assumption

(Z) $\vdash E_r^a \equiv E_s^a$,

as will be the case if, for example, s is defined as $2r$ minus the frequency of a.

Then we *cannot* derive from (Z), either alone, or in combination with (Rc) or (Rcr) that

$$r = s.$$

For assume that $r = 2$ and $s = 3$. In this case, E_r^a and E_s^a will both be self-contradictory (because no relative frequency or probability can exceed 1); and as a consequence, (Z) will hold, since any two self-contradictory statements are inter-deducible. Yet in spite of (Z) we cannot deduce $r = s$. Moreover, we *cannot do so* by applying the following method: we first derive:

(1) If E_r^a is true, then $P(A) = r$ (by Rc)

and

(2) If E_s^a is true, then $P(A) = s$ (by Rc)

and we now apply (Z) to (1) and (2). For applying (Z) to (1) and (2) does not allow us to derive from (1) and (2) respectively

(—) $P(A) = r$

(—) $P(A) = s$

from which indeed '$r = s$' would follow; rather, in order to derive these two equations (—), we would have to have 'E_r^a is true' and 'E_s^a is true', and so obtain the equations (—) from (1) and (2) by the rule of detachment.

---

[1] That these rules are, in fact, not acceptable may be seen from my paper 'Probability Magic, or Knowledge out of Ignorance', *Dialectica*, 43–44, 1957, pp. 354 ff. (to be republished in a forthcoming collection of essays of mine).

All this changes completely if we express the 'straight rule' which we have tried to formulate by (Rc) or (Rcr) by the following rule (R) which to some people may seem intuitively just as convincing.

$$\text{(R)} \qquad P(A, E_r^a) = r,$$

and which may be read

'The probability (or degree of confirmation) of A, given as evidence E_r^a, equals r.'

This new rule (R) looks very much like our previous rules. However, if we assume that the general probability calculus applies to our P-expressions, that is to say, to all expressions of the form

$$P(A, B),$$

then we obtain from (R) and (Z) at once the paradoxical conclusion

$$r = s$$

For the calculus of probability contains the universal rule,

$$\text{(Z')} \qquad \text{If} \vdash B \equiv C, \text{ then } P(A, B) = P(A, C).$$

Admittedly, in most formulations of the calculus of probability, (Z') is valid only under the assumption that $P(B) \neq 0 \neq P(C)$; and this might seem to offer a way out of the difficulty. However, I have shown (in my *Logic of Scientific Discovery*) that a consistent calculus of probability can be constructed in which (Z') is universally valid *without* the assumption that $P(B) \neq 0 \neq P(C)$. Moreover, I wish at this stage of my discussion only to draw attention to the fact that (R) may entail, in spite of its intuitive similarity to (Rc) or (Rcr), conclusions which cannot be reached from (Rc) or (Rcr), and that the apparent intuitive similarity between (R) on the one side and (Rc) and (Rcr) on the other side should therefore be regarded with suspicion: it may turn out to be highly misleading.

The situation is simply this: if we wish to derive from

$$\text{(R)} \qquad P(A, E_r^a) = r$$

from which we obtain by substituting 's' for 'r'

$$\text{(R)} \qquad P(A, E_s^a) = s \qquad \text{(by R)}$$

the conclusion

$$r = s$$

then we need not more than (Z); while for the corresponding derivation from (Rc) or (Rcr), *more than* (Z) *is needed:* we also need

E_r^a is true, and E_s^a is true.

This shows that there is a great difference between (R) on the one hand and (Rc) or (Rcr) on the other.

I now proceed to show formally how we can derive a contradiction from rule (R). I shall strengthen the straight rule (R) in what follows by the condition '$0 < r < 1$'.

II

We assume the apparently intuitively satisfactory form of the '*Straight Rule*':

(i) If $0 < r < 1$, then

$$P(A, E_r^a) = r$$

where 'E_r^a' is the name of the statement 'the observed frequency of a equals r'; or, written in symbols, 'E_r^a' is the name of the equation:

$$fr(a) = r.$$

We obtain from (i), by substituting 's' for 'r',

(ii) If $0 < s < 1$, then

$$P(A, E_s^a) = s.$$

We also have identically for any x whatever:

(iii) If $\vdash s = 2r - x$, then the following equivalence holds analytically:

$$x = r \text{ if and only if } x = s.$$

By substituting '$fr(a)$' for 'x' we obtain from (iii)

(iv) If $\vdash s = 2r - fr(a)$, then the following equivalence holds analytically:

$$fr(a) = r \text{ if and only if } fr(a) = s.$$

By using the metalinguistic names 'E_r^a' and 'E_s^a', this can be written:

(v) If $0 < r < 1$ and $0 < s < 1$ and $0 < fr(a) < 1$, and if $\vdash s = 2r - fr(a)$, then

$$\vdash E_r^a \equiv E_s^a.$$

But (v) allows us (like (Z′) above) to substitute 'E_r^a' and 'E_s^a' in every probability formula for each other. Thus we obtain:

(vi) If $0 < r < 1$ and $0 < s < 1$ and $0 < fr(a) < 1$, and if $\vdash s = 2r - fr(a)$, then

$$P(A, E_r^a) = P(A, E_s^a) = r = s$$

and therefore

(vii) If $0 < r < 1$; $0 < s < 1$; and $0 < fr(a) < 1$, and if $\vdash s = 2r - fr(a)$, then

$$r = s.$$

But this result, it can be seen at once, is inconsistent: for first of all, we may *for any given number*, $fr(a)$, such that $0 < fr(a) < 1$, choose an r and an s such that r is not equal to s even though the conditions of (vii) are satisfied. For example, for $fr(a) = 0$ we may choose $r = 1/2$ and $s = 1$; and for any given $fr(a)$ such that $0 < fr(a) < 1$ we may choose $r = fr(a)/2 > 0$ and $s = 0$.

In all these cases the conditions of (vii) are satisfied and so we arrive at a straightforward contradiction.

But we can also satisfy the conditions of (vii) by substituting '$2r - fr(a)$' for 's' in (vii). This leads to:

(viii) If $0 < r < 1$ and $0 < fr(a) < 1$, then $r = 2r - fr(a) = fr(a)$.

III

The preceding derivation of (viii), i.e. of

$$fr(a) = r$$

(which is clearly contradictory, since r and $fr(a)$ are free variables, so for any given $fr(a)$ we can choose an r such that '$r \neq fr(a)$' holds) can be much abbreviated. If we permit in the step from (i) to (ii) the substitution of '$2r - fr(a)$' instead of 's' for 'r', we may arrive at:

(ii′) If $0 < 2r - fr(a) < 1$, then

$$P(A, E^a_{2r-fr(a)}) = 2r - fr(a).$$

Then, on the basis of the logical equivalence of the two equations

$$fr(a) = r$$

and

$$fr(a) = 2r - fr(a)$$

we obtain

(iii′) $\vdash E^a_r \equiv E^a_{2r-fr(a)}$

which leads directly to the inconsistent formula

(iv′) If $0 < r < 1$ and $0 < fr(a) < 1$, then

$$fr(a) = r$$

which is the same as (viii).

I do not doubt that the substitutions on which this derivation depends are incontestable; but since the attempt might be made of introducing special *ad hoc* rules of substitution forbidding the substitution of expressions containing '$fr(a)$' in 'E'-symbols, I have given above my derivation (i) to (vii) and (viii) in which no such substitutions are made.

IV

In the rest of this note I shall follow Miller's own method as closely as possible in order to show its applicability to Carnap's system of inductive methods. The reason for following Miller's method so closely is that it seems to suggest, I think, a (somewhat desperate) way of rescuing the situation of the inductivists. I shall discuss this at the end of my note.

I shall in what follows not refer back to the formulae that I have so far used (they are distinguished by their roman numerals).

In order to apply Miller's result to Carnap's Theory of Induction it is necessary to strengthen slightly Carnap's *metalinguistic symbolism*. (There is no need to strengthen Carnap's object language.)

Carnap uses 'h_M' as a metalinguistic name of the *hypothesis* that some individual thing (the subject of h_M) not included in the observed sample, has the property M.

I shall write 'H^M' in place of Carnap's 'h_M' for typographical reasons which need not be explained.

Carnap uses 'e_M' as a metalinguistic name of the total available *evidence* pertaining to the s observed things that have or have not the property M. This evidence statement e_M is supposed to consist of a conjunction of s different observation statements, about a given set of s different individuals, where the number s is called the *sample size*. Each of these observation statements states of one of the s individual things of the given observed sample whether or not this individual thing has the property M.

The user of the metalanguage is supposed to be able to know of each e what its sample of individuals is, and to ascertain (by counting) the sample size s and the number s_M of those conjunctive components of the conjunction e_M which affirm of their individual subject that it has the property M. Or in other words, the user of the metalanguage is supposed to be able, by scrutinising or analysing a certain evidence statement e_M, to discover that it refers to s different individuals, and that it asserts of s_M of these s individuals that they have the property M. (Thus we have $s_M < s$.)

Carnap's own metalinguistic name 'e_M' does not disclose some of the most important properties of the evidence statement e_M which it denotes: it mentions that e pertains to M, but it omits to indicate the important fact that e contains s subjects of which s_M have the property M.

I propose to remedy this defect of Carnap's metalinguistic notation by introducing, instead of his symbol 'e_M', a new symbol

$$E^M_{s_M,\, s_N}$$

as a more fully descriptive name of the total evidence. Here 'e_M' is replaced, for typographical reasons, by 'E^M'; s_M is the number of individuals (mentioned in the evidence) having the property M; and s_N is defined by

Definition $\qquad s_N = s - s_M$.

Thus s_N is the number of those of the individuals (mentioned in the evidence) which do not have the property M; and I shall use altogether (for reasons of typographical simplicity) 'N' as the name of the complement of M.

It should be noted that s, s_M and s_N are all numerical variables, and that I use the name

$$E^M_{a,\, b}$$

so as to imply that $s_M = a$ and $s_N = b$, so that $s = a+b$. In other words, the *first* of the two numerical *subscripts* gives the number of those individuals which have the property mentioned by *the raised property symbol.* Thus

$$E^A_{a,\,b}$$

would imply that $s_A = a$ and $s = a+b$;
and

$$E^N_{s_M,\ s_N}$$

would imply that $s_N = s_M$, because s_M is here the *first* of the two numerical subscripts and therefore belongs to the property mentioned by the raised property symbol, which in this case is 'N'.

It follows from these explanations, and from Carnap's own condition C 7,[1] that the following principle must hold:

(a) $E^M_{a,\,a}$ is substitutionally equivalent (in all relevant *c*-formulae) to

$$E^M_{s_N,\ a}$$

since each of these two metalinguistic names describes essentially the same evidence situation: both say that the numbers s_M and s_N are equal, so that

$$s = 2s_M = 2s_N.$$

Like Carnap, I write 'c' rather than 'p'; so '$c(H^M, E^M_{a,\,b})$' denotes his 'degree of confirmation of H_M, given the information or evidence $E^M_{a,\,b}$'.

We can now proceed with Miller's argument, as follows:

We first assume exactly as he does that the true experimental evidence is based on throws of a die with the property M stating, say, that the number 5 turns up:

(α) $E^M_{1,\,5}$ is true.

This implies that we have made six individual throws, so that $s = 6$, $s_M = 1$ and $s_N = 5$.

The assumption (α) will not, however, be used in any way—except as a reminder, at the very end of our derivation, how very absurd the result is that will be derived there (in 7).

We begin by formulating the 'straight rule' as follows:

(1) $$c(H^M, E^M_{s_M,\,s_N}) = s_M/s \qquad (s = s_M+s_N)$$

or what amounts to the same

(1′) $$c(H^M, E^M_{a,\,b}) = a/(a+b).$$

We obtain by substitution

(2) $$c(H^M, E^M_{3,\,3}) = 3/6 = 1/2 \qquad \text{(by 1′)}$$

and also

(3) $$c(H^M, E^M_{s_N,\,3}) = s_N/(s_N+3). \qquad \text{(by 1′)}$$

---

[1] See in this connection Carnap's condition C 7 on p. 14 of his book *The Continuum of Inductive Methods*, 1952. Condition C 7 states, in essence, that only the *numbers* of the different individuals mentioned in *e* are relevant, not their individual names.

But by (a) we have also

(4) $E^M_{3,3}$ is substitutionally equivalent in all relevant c-formulae to $E^M_{s_N,3}$. (by a)

Accordingly we can obtain

(5) $$1/2 = s_N/(s_N+3)$$ (by 2, 3, 4)

and therefore

(6) $$s_N = 3 = s_M.$$ (by 5)

This result is, of course, absurd, since it is informing us of an empirical situation although we have assumed nothing but the formal rule (1); or in other words, it is absurd because we have obtained empirical information from a purely formal rule. This is illustrated by the fact that we have assumed, according to (α), that the empirical facts are otherwise. That is, according to (α), $s_N = 5$. So we have

(7) $$3 = 5.$$ (by 6, α)

which is the same as Miller's contradiction.

By the same method contradictions can be derived for *all* of Carnap's non-extreme methods in his *Continuum of Inductive Methods*. For example, we obtain from his (15-3), p. 45, for our die and the property M the general rule

(1*) $$c^\star(H^M, E^M_{a,b}) = (a+1)/(a+b+6).$$

By making the same substitutions for a and b as before, we obtain

(5*) $$4/12 = 1/3 = (s_N+1)/(s_N+3+6)$$

and therefore as before

(6*) $$s_N = 3$$ (by 5*)

(7*) $$3 = 5.$$ (by 6*, α)

Miller's method of constructing a contradiction is applicable to all of Carnap's methods except *one*—the 'first extreme method', as he calls it (*Continuum*, p. 37); but this is *not a method of induction*, since then Carnap's 'degree of confirmation' is given *a priori*, and remains *completely unaffected by any experience*.

There seems to be only one hopeful way out of all this for probabilistic induction. It would amount to an abandonment, or at least a severe modification, of the idea of a *relative* (or *conditional*) *inductive probability*.

For we may try to avoid Miller's paradox by writing (1′), as indicated in section I, as a *conditional straight rule* as follows:

(Rcr) If $E^M_{a,b}$ is true, then $c(H^M_{a,b}) = a/(a+b)$.

But if we do this, there does not seem—in the light of the probabilistic rule of detachment—any real need for the *relative* probability, $c(x, y)$; y becomes in fact redundant, so that we can write Reichenbach's straight rule:

(Rc) If $E^M_{a,b}$ is true, then $c(H^M) = a/(a+b)$.

A reform on these lines turns out to be free of Miller's contradiction (though not of those referred to in my second footnote).[1] But it would entail the *complete abandonment of the philosophical idea that there is an 'inductive logic'* which is, formally, identical with the calculus of relative probabilities—that is, with the *probabilistic generalisation of deductive logic.*

London University KARL R. POPPER

Functional Analyses in Biology

1. John Canfield's views concerning functional analyses in biology have an appealing plausibility, but we believe that they are defective both in detail and in conception.[2] In particular, we object to the translation schema which he proposes for functional analyses:

> A function of I (in S) is to do C *means* I does C; and if, *ceteris paribus*, C were not done in an S, then the probability of that S surviving or having descendents would be smaller than the probability of an S in which C is done surviving or having descendents (p. 292).

In Canfield's opinion, a functional analysis of an organ is an attempt to state what the organ does that is useful to the organism of which it is a part. His translation schema develops out of the following preliminary account of functional analyses:

> FA is a correct functional analysis of I *if and only if* FA has the linguistic form of a functional analysis, and FA states what I does in S, and what I does is useful to S (p. 288).

Given this preliminary account, he readily derives his translation schema by elucidating the concept of 'useful'.

According to Canfield, there are two ways in which a biological item may be useful: 'either what the item does is useful by aiding to preserve the life of things having the item, or else what it does is useful by aiding to preserve the species of things having it' (p. 291). In this translation schema, usefulness for individual survival is explicated in terms of the probabilities that an individual will survive with and without the item in question; usefulness for species survival is explicated in terms of the probabilities that an individual will have descendents with and without the item in question.

---

[1] (Added in proofs.) I have found since that the reform here indicated runs into serious trouble and that it is not effective in avoiding Miller's paradox. For this a more radical reform would be required, and one that is obviously *ad hoc.* (This should not be interpreted as suggestion that this more radical reform might not, in its turn, run into some similar trouble.) I intend to explain this—the breakdown of the rules (Rcr) and (Rc)—in a further note.

[2] John Canfield, 'Teleological Explanation in Biology', this *Journal*, 1964, pp. 285-295; and 'Teleological Explanation in Biology: A Reply', ibid. 1965, pp. 327-331.

2. Now it is not correct that the contribution of an item to species survival can be evaluated by comparing the probability that individuals with the item will have descendents and the probability that individuals without it will have descendents. First, a species will not survive unless its members' descendents are *fertile*. Second, the *number* of descendents is clearly relevant to species survival. For these reasons, it would appear to be in Canfield's own interest to revise his translation schema in something like the following way:

> A function of I (in S) is to do C *means* I does C; and if, *ceteris paribus*, C were not done in an S, then either the probability of that S surviving would be smaller than the probability of an S in which C is done surviving, or the probability of that S having a number of fertile descendents within the optimal range would be smaller than the probability of an S in which C is done having a number of fertile descendents within the optimal range.

It is necessary to put the matter in terms of a number of fertile descendents within the optimal range, rather than simply in terms of a greater number of fertile descendents, because it cannot be assumed that a species is the more likely to survive as its members have a larger number of fertile descendents. Allowance must be made for the possibility that a 'population explosion' would actually work against the survival of the species.

3. Canfield claims that the acceptability of his translation schema is supported by 'the seeming correctness' of his way of understanding 'useful' (p. 293). But even if his account of usefulness is improved by making the revision we have suggested, what reason is there to think that it is correct? The most relevant sense in which it might be correct is that of conforming to the usage of biologists who offer functional analyses. However, Canfield presents no evidence that biologists understand usefulness in the way that he defines.

To be sure, his approach to the question of how an item may be useful in species survival recalls the familiar notions of Darwinian fitness, selective advantage, and adaptive value, which are widely employed in evolutionary studies. But he associates this concept of usefulness with another which is not of equal standing in discussions of evolution. What helps an individual to survive does not necessarily have adaptive value, confer selective advantage or contribute to Darwinian fitness. A process which is useful to the survival of a species may well be hazardous to its members, the most obvious example being childbirth.

Naturally, a species cannot survive unless its members live long enough to reproduce and to provide whatever care their offspring require. But in so far as it is relevant, reference to individual viability is included in the notion of a species' fitness; and the latter is understood in terms of the ability to produce fertile offspring. The ability of individuals to survive is not, as in Canfield's account, a disjunctive element in the criterion of fitness which evolutionary theory employs. Thus, his conception of usefulness differs from that invoked by those biologists whose approach to the analysis of function might appear to be closest to his own.

4. To the extent that his conception of usefulness corresponds to notions which are integral to evolutionary theory, Canfield seems committed to maintaining that the functional analyses accepted by contemporary biologists differ in meaning from those which biologists accepted before the theory of evolution developed. This is possible, but it is not very plausible. The concept of biological function is an old one, and numerous standard descriptions of functions do not appear to have changed in meaning with the advent of evolutionary theory.

5. To decide whether performing a certain activity is a function of an organ, Canfield would have us compare a specimen in which the activity is performed with one in which it is not performed. Of course, the comparison will not be significant unless the two specimens are sufficiently alike. Canfield attempts to provide for this condition by including a *ceteris paribus* clause in his translation schema.

By *ceteris paribus* he means, among other things, that the two specimens are alike in all relevant respects other than the preformance of the activity in question. But if a certain activity is performed in one specimen and not in another, it is impossible that the two specimens should differ only in this. An organism's activities arise out of its bodily structure, and its structure must be altered in order to change its activities. The notion of two organisms which have the same structure, but in which different activities take place, violates our ideas of causality.

Suppose we wish to know whether secreting bile is a function of the liver in mice. Canfield advises us to examine a mouse which does not secrete bile but which is otherwise identical with a mouse whose liver does secrete it. There can be no such mouse. If it were built the same as other mice, it would naturally secrete bile just as other mice do; to prevent it from doing so, we would have to give it a peculiar sort of liver and, doubtless, other aberrant characteristics as well. Moreover, there are an infinitude of possible mouse-like organisms which do not secrete bile. Which of these does Canfield propose that we compare with the bile-producing specimen?

Similarly, consider what Canfield says about whether it is a function of the heart to produce a pulse:

> Suppose that we could add to an animal's heart an instrument which resulted only in making the animal's blood flow at a steady rate, rather than in spurts. So far as we know this change would neither raise nor lower the animal's chances of survival (p. 330).

Surely it is obvious that an animal with such an instrument added to its heart does not differ from other animals of the same species only in lacking a pulse? Unlike the others, after all, it has an instrument added to its heart. One cannot merely eliminate an animal's pulse and see what happens; one must alter the animal's structure in order to deprive it of a pulse. There are any number of such alterations which might be made. Moreover, since there is not a neat one-to-one correspondence between structures and activities, these alterations must be presumed to have additional effects besides that of eliminating the pulse. We find it difficult to share Canfield's confidence that an animal with an instrument added to its heart has the same chance of survival as one without it.

We find Canfield's use of *ceteris paribus* to be biologically meaningless. As it appears in his translation schema, the phrase is an empty locution whose pseudo-experimental flavour conceals a fundamental difficulty in his approach to the understanding of functional analyses.

6. In any case, Canfield's translation schema gives rise to a view which he himself explicitly rejects. He regards it as false that a function of the heartbeat in vertebrates is to produce heart sounds; indeed, he takes this to be a paradigm of false functional analysis. However, if his translation schema is correct, it is true that a function of the heartbeat in vertebrates is to produce heart sounds.

The usefulness of a biological item, in Canfield's sense, depends partly upon the environment of the organism in which the item occurs. Now the present environments of many vertebrates include physicians, and the practice of physicians involves making diagnoses which often rely on the character of their patients' heart sounds. The patients presumably have better chances of surviving and of reproducing if this diagnostic technique can be used on them than if, *ceteris paribus*, their health could not be evaluated by listening to the sounds which their hearts make. On Canfield's view, then, it is true that a function of the heartbeat in vertebrates is to produce heart sounds. We would be interested to know whether he is prepared to follow his account of functional analyses to this conclusion.

7. While we have no alternative translation schema to offer, we may make the following suggestion. Canfield's account derives its plausibility from the fact that the notions of survival value, selective advantage and fitness are indeed relevant to the understanding of organic functions. But the relevance of these notions is that they enable us to understand why certain functions are performed. One may appreciate this without supposing that the notions are constituents of the meanings of functional analyses.

The Rockefeller University

Harry G. Frankfurt
Brian Poole

REVIEWS

The Discovery of Time. By Stephen Toulmin and June Goodfield. Hutchinson, London, 1965. Pp. 280. 35s.

Both the strengths and the weaknesses of this book derive from the authors' intrepidity. They had to be dauntless to attempt it: the aim is nothing less than 'to depict the gradual emergence of a continuing sense of history out of earlier mythological and theological systems' (p. 15). In attempting to answer the question 'How did our intellectual picture of Nature acquire its temporal perspective?' they have taken responsibility for reviewing just under five thousand years of the history of ideas on the terrestrial, animal, human and cosmic past. Thus, the book includes chapters on Greek and Christian philosophy, the Scientific Revolution, the development of civil history, the history of geology and biology. Evolution, of course, is the dominant theme. The work under review is only part of the overall task which they set themselves: 'to trace out the main lines of intellectual development by which the fundamental features of our contemporary picture of Nature came to be accepted as common knowledge' (p. 15). Three volumes have now appeared, and a fourth is to follow.[1]

This series—and particularly the Authors' Foreword to the latest volume raises three issues which should be mentioned before presenting a criticism of the text. First, the authors have indicted the established authorities and institutions in Britain for their failure to follow up the work admirably begun by the Nuffield Foundation. We are informed that 'As things stand, it appears that educational foundations are the only real means by which novel, exciting—even eccentric—academic ideas can get off the ground' (p. 16). It is implied that narrowmindedness is responsible for the authors' emigration. Second, our attention is drawn to the fact that 'in the British academic world today the history of ideas still remains (to use Berlin's pungent phrase) " a non-subject " '. The authors hope that their 'attempts to reintegrate the ideas of the natural and human sciences may have done something to awaken a greater interest in this field of thought' (ibid.). Third, the style of the book, its availability in a schools edition at 15s., and the reaction of the lay press raise the issue of popularisation. Whatever professional historians and philosophers may think of this book, its reception implies a demand for such works which places an obligation on the academic community. According to a publisher's blurb, the books in this series have been hailed as 'original', 'dashing and brilliant', 'absorbing . . . a rare achievement'. If we are not satisfied to have our research represented by this series, we shall just have to write better books ourselves.

---

[1] *The Fabric of the Heavens*, 1961; *The Architecture of Matter*, 1962. Both have been reprinted by Penguin Books. The forthcoming volume is *Science and Its Environment*.

We are left with a picture of a tradition-bound academic community which is hostile to intellectual history and indifferent to the interests of the general reader. There is much justice in these claims. British universities are inordinately cautious about fostering new subjects, and those who are competent to do so have perhaps been too reluctant to make the results of specialised research available to non-specialists. Nevertheless, it should be pointed out that the case against the history of ideas rests on the fact that few scholars besides Lovejoy have lived up to the standards which he set. Instead of becoming the most challenging of intellectual disciplines, 'history of ideas' has too often served as an excuse for not doing one's homework. In order to trace the history of an idea one must be very erudite indeed, and in order to popularise history one must prepare a highly distilled summary, based on a wide reading of primary sources in the light of the best specialised scholarship. In attempting popularisation and the approach of the historian of ideas at the same time, Toulmin and Goodfield have here produced a book which does nothing to mitigate one's reservations about 'history of ideas'. On the contrary.

Before going further with the extremely unpleasant task of writing a severely critical review, I should stress that the authors write interestingly and ask incisive questions about their materials. In particular, they have an ability to identify what was left out—the questions which were not asked in a given period. There are also many useful (if sometimes facile) contrasts, e.g. 'Lamarck's difficulty was to produce a convincing explanation of organic *change*. Cuvier's problem was the reverse: if there were renewed organic creations after every catastrophe, that accounted for the difference between the animals of different epochs, at the cost of leaving their *resemblances* mysterious' (p. 183).

This brings us to the second aspect of the authors' intrepidity. Literally the term means 'not alarmed'. In this case they should have been, since the main point about this book is its appalling standard of scholarship. After making due allowance for the immense task which the authors set themselves it remains the case that the book is nowhere near good enough either as intellectual history or as popularisation. As a combination of the two it is superficial and misleading. One could make a long list of omissions, but the aims of the book make this unfair. One could list disagreements about emphasis (e.g. the problem of Being *versus* Becoming in Plato and Aristotle is not made sufficiently clear, and its persistence as a theme in the history of genetic thinking is insufficiently emphasised), but the same defence might be offered. On matters of accuracy one is in a difficulty because of the vast scope of the book. M. I. Finley has been very critical of the treatment of the Classical period,[1] and in my own period, I find it less than reassuring to find Darwin's views in 1859 quoted and then contrasted with those of 1872, with *all* quotations taken from the sixth (1872) edition of *The Origin of Species.*

---

[1] *New Statesman* for 11th June 1965, pp. 925-926.

And then there are simple gaucheries, e.g. we are told in an aside that the 'hot-line' telephone between Moscow and Washington is 'the first tangible element of a new world-wide political system' (p. 234). But these piecemeal remarks do not add to up one's main criticism. The book is full of interesting information, but the authors have not transformed their materials in the light of their general topic. The central theme is not central, and their materials remain parts of disparate studies. The chapters give no more sense of a coherent thesis than the chapter headings. That is, *The Discovery of Time* is not the history of an idea.

The mode of presentation seems designed to preclude further study. The omniscient 'story-telling' style obscures the fact that the book is about issues which we are still far from understanding. Historical scholarship consists of controversy, not settled stories. Consequently, an introductory work should raise questions, point to loose ends, and lead the reader to look further. Worse still, the book contains no footnotes, the references are incomplete, and few of these are to primary sources. While it is too authoritarian in its presentation, it lacks authority in its substance. One often feels that the authors are telling all they know on a given topic and that we are being provided with summaries of secondary and tertiary sources written about works which have not themselves been carefully studied. This induces a regrettable distrust in the reader.

Having said this, I should add that there is one hopeful sign that the authors have identified a very important idea as a result of their studies which should bear fruit, no matter how disappointing the present volume is. The most exciting issue raised in the book concerns the future of science: 'The outstanding question now is, whether the laws of Nature themselves—the last *a-historical* feature of the physicists' world-picture—will in their turn prove to be subject to the flux of time' (p. 250). 'Do all the fundamental laws and constants of physical theory retain their forms and values eternally, without any modification? About this we can no longer be certain. We may yet be on the threshold of the greatest of all scientific revolutions' (p. 265). It is hoped that Toulmin and Goodfield will tell us more about this idea and perhaps investigate other ways in which evolution and natural selection are becoming paradigmatic for modern views of science and history. This interpretation has been applied to the methods and the development of science by Popper, Feyerabend, Kuhn, and (elsewhere) Toulmin himself. It forms the basis of modern learning theory in psychology and in some of its liaisons with cybernetics. C. S. Peirce was perhaps the most daring exponent of the view that both the elements and the laws of nature are constantly evolving. Here is an idea which is perhaps more important to history, philosophy and science than any other since Plato sacrificed time and change for the sake of certainty.

In conclusion, the failure of this book as a study in the history of ideas may serve as an appropriate occasion for repeating Lovejoy's *desiderata* for a history of genetic thinking. They were written in 1938 and, unfortunately, remain unfulfilled.

'To turn to the history of (primarily) scientific conceptions: there as yet exists, so far as I know, no historically and philosophically respectable account of the total development of the idea of evolution before Darwin—using the term even in its narrower sense of the theory of the transformation of species; and we have certainly no adequate history of the idea in its broader sense, i.e. of developmental conceptions in astronomy, geology, anthropology, social philosophy, cosmology and theology, and the influence of all these in other provinces of thought. Historically, the various phases of the progress of what may be called the genetic way of thinking—which has been a long, complex, and extremely gradual process—are closely related. The reason why there exists no adequate history of the process as a whole is, in part, that much of the *grundlegend* detailed study of sources still remain to be done; but the task, in any case, can scarcely be executed properly by one scholar. For it requires a competent acquaintance with many special fields—not only with the several natural sciences mentioned, and with the history of particular subdivisions of these—in biology, for example, of taxonomy, comparative anatomy, paleontology, embryology and genetics—but also with an extremely wide range of the philosophical, theological and general literature of the seventeenth, eighteenth, and early nineteenth centuries, and with the history of ideas *about* history and its general movement. A part of the story, for example, can be verified only through a thorough study of the works of Leibniz; and another part demands an intimate acquaintance with the literature, the metaphysics, and even the aesthetic theories and fashions, of the German Romantic period. And all of these need to be illuminated from the special history of certain relevant individual ideas—for instance, of the principle of continuity and of the concept or pseudo-concept of " species ". The thing can be done, it could be wished that American scholarship would undertake it. But unless it is undertaken in the manner suggested, through large-scale, planned, and articulated co-operation, it is likely to be much more imperfectly done than even the present state of knowledge in the numerous special fields pertinent to it would make possible.'[1]

ROBERT M. YOUNG

Beyond the Edge of Certainty: Essays in Contemporary Science and Philosophy.
R. G. Colodny, Ed.
Englewood Cliffs, N.J., 1965. Pp. 287. 68s.

THE essays in this volume, the second in the Pittsburgh Series in the Philosophy of Science, are as uneven as its title is ambiguous. As such collections go, however, this one offers a rich and relatively varied cornucopia. Particularly impressive is the effort all the contributors made to cross disciplinary lines in order to say something significant to the scientist, the philosopher and the historian of science. The least successful essays, from this point of view, are Morrison's 'The Physics of the Large', which recapitulates and oversimplifies recent work in cosmology, and Rescher's 'The Ethical Dimension of Scientific Research', which makes some obvious observations about the morality of what Derek Price has called 'Big Science'. But except for these two selections, the volume could be profitably read by all those interested in the philosophical problems of classical and contemporary physics. The following remarks are intended only to provide vignettes of the papers in the book.

---

[1] 'The Historiography of Ideas', reprinted in *Essays in the History of Ideas*, Johns Hopkins, Baltimore, 1948, Capricorn Books edition, New York, 1960, pp. 12-13.

With the flair and verve we have come to expect from him, Professor Hanson ('Newton's First Law: A Philosopher's Door Into Natural Philosophy') tackles Newton's first law of motion in order to show that even apparently simple cases in the history of science are pregnant with problems for the philosopher. Taking on the rôle of midwife himself, Hanson proceeds to argue that the principle of inertia, though ostensibly a single law, is really an algorithm for a wide variety of laws. Depending upon whether we take 'force-free', 'rectilinearity', 'uniform' or 'motion *ad infinitum*' as the primitive term, we may get radically different conceptual schemes in physics. Hanson explicates the semantic linkages between terms like 'uniform' and 'force-free' and uses these linkages to argue for the non-falsifiability of the law of inertia. He thus reminds the philosopher of science that most of science's interesting philosophical problem occur in *loci classici*—Galileo, Newton, Lagrange and Hertz—as well as in contemporary physics. To those of us who have difficulty seeing the philosophical forest for the mathematical trees in modern science, his point is cheerfully embraced. Though Hanson's general thesis (viz. that the history of physics offers fascinating case-studies to the philosopher) is well-taken, his argument is handicapped by one of the examples he uses. Specifically, Hanson claims that Galileo's law of inertia attributed rectilinear motion to moving bodies acted upon by no forces. He neglects the fact that Galilean inertial motion was circular; from this, no amount of semantic unpacking could yield 'rectilinearity' as a property of Galilean force-free motion. While circular motion does approximate rectilinear motion in terrestrial cases, there still remains a vast conceptual difference between the two which should not be so glibly glossed over. Hanson's reluctance to do his homework here highlights his own plea for the need for philosophers to study history. It should also be noted that the essay largely repeats an earlier article and assumes both the insights and the prejudices of the previous attempt.

Ellis's contribution, 'The Origin and Nature of Newton's Laws of Motion', makes Hanson's point as to the philosophical value of history by example rather than by proselytising precept. Ellis sets out to determine the grounds on which a law of motion could be rationally accepted or rejected. Taking his cue from Duhem, Ellis shows by a careful historical analysis that Newton's laws of motion were neither derived from, nor directly based on, any experimental evidence, notwithstanding Newton's inductivist remarks to the contrary. He suggests (as Dijksterhuis has also recently suggested) that Newton's force law cannot be unambiguously rendered as F=ma. The force law of the *Principia* is actually $\Delta F = \Delta$ (mv) (i.e. change of impressed force is proportional to change of motion or momentum). The force law, in Newton's hand, made no allusions to the time-rate of change of velocity. But Ellis's hatchet-work does not stop here. He goes on to undermine the traditional view that Newton's laws derived from Galileo's study of free-fall and suggests, rather, that Newton's dynamics stems from the work of Descartes and Huygens on impact. Moving with ease from historical (but not for that reason antiquarian) puzzles such as these, to strictly philosophical questions about the nature of the laws of motion, Ellis convincingly argues that the law of inertia is not an empirical proposition but rather a principle which defines what types of motion (uniform and rectilinear) need no explanation, i.e. what types of motion do not require the invocation of forces for their explanation. Since we only attribute forces to situations in which bodies move unnaturally (viz. non-inertially), the existence of forces is logically contingent upon our selection of what is to constitute 'natural' motion. Ellis argues, further, that philosophers are mistaken when they talk about *the* status of the force law in scientific theories because such talk presupposes that this law plays only one role in physics. On the contrary, in some contexts it functions as a definition, while in others it specifies empirically observable relations between force, mass and acceleration. Formal analysis of scientific concepts is thus handicapped by its neglect of the contexts in which scientific statements are uttered and used.

In the next essay ('A Philosopher Looks at Quantum Mechanics'), Putnam critically examines several alternative formulations of the quantum theory (De Broglie's, Born's, Bohm's and Heisenberg's). He argues that the Copenhagen interpretation poses as many thorny philosophical problems and leaves unresolved as many anomalies as earlier interpretations. In particular, Putnam insists that Heisenberg *et al.* have been unable to explain, except in a pathetically *ad hoc* way, why macro-observables always have definite numerical values when their micro-observable components do not. His conclusion, which many philosophers undoubtedly share, is that there exists as yet no satisfactory interpretation for the formalisms of quantum mechanics.

In an essay that allegedly 'restores the legitimacy of teleological concerns', David Hawkins ('The Thermodynamics of Purpose') dabbles with a variety of themes which are designed to show that modern science acknowledges the existence of purposes and that the mechanical philosophy, far from eliminating teleological language, has introduced a teleology of its own.

The author cites the following types of teleology in physics: (1) the teleology of mechanistic laws, (2) the teleology of quantum states, and (3) the teleology of thermodynamic processes, especially those relating to information theory. In (1) Hawkins aligns himself, perhaps unconsciously, with Robert Boyle, who likewise argued that the mechanistic laws of physics function as final causes. But what was permissible philosophy in the seventeenth century is not necessarily sound today. Ultimately, Hawkins's argument rests on a pun with the word *telos*, and a subtle assimilation of Aristotelianism to modern science. Aristotle's *telos*, we are told, was a form or nature. Mechanistic laws can be viewed as forms or descriptions of the nature of physical processes. Thus, mechanistic laws express the *telos* of matter and are, to that extent, teleological. Modern scientific explanation, in terms of mechanisms and models, 'is just the sort of thing Aristotle talked about under the heading of ends'. (104). Never mind, Hawkins pleads, the anthropocentric overtones in Aristotle; they were an accident of his view of nature and are not essential to his teleology. On the contrary, however, if we divest Aristotle's final causality of its homocentric trappings, nothing of interest is left; for it is the anthropocentricity which makes Aristotle's view teleological. The Aristotelian concept of substance, on which his account of physical causality is based, rests on human action as the paradigm case of change (*kinesis*). If you throw out the anthropocentric bias, then Aristotle's doctrine of *informed* matter becomes meaningless. In (2) and (3), Hawkins's arguments are much more plausible. Even so, given the irritating ambiguity in his use of 'teleology', it is very difficult to see that they amount to anything more than the innocent claim that, as he puts it, 'physical effects are not always adequately explained by physical causes' (114).

With little doubt, the most exciting and fundamental essay in this volume is Feyerabend's 'Problems of Empiricism.'[1] Though ostensibly concerned with the philosophy of science, the discussion ranges knowledgeably over psychology, perception, epistemology and the theory of meaning. Fortunately, the analysis is as careful as its scope is broad. Feyerabend argues against the monistic empirical view that it is a mistake to accept mutually inconsistent theories. Empiricists, in their belief that a scientific theory is a putatively true description of nature, deny that two contradictory theories can be used simultaneously because at least one of them must be false. Since science is conceived as a search for the real and the true, only one theory can be entertained at a time, and it will be believed until it is falsified when another unitary theory will take its place. In contrast to this position, Feyerabend seeks to formulate an empiricist philosophy which allows room for the interplay of alternative theories and conceptual schemes;

---

[1] However, it must be confessed that one's enthusiasm is slightly diminished by discovering that this essay is an enlargement of Feyerabend's paper in the second volume of the Delaware Series in the Philosophy of Science which was, in turn, a revision of a paper in *Inquiry*.

in short, for a theoretical pluralism. He shows that the falsifiability of a theory T is increased when we have alternative theories, $T_1, T_2, \ldots, T_n$, which will direct our experiments along lines of research, relevant to T, that T itself would never have suggested. Generalising the work of Hanson and Duhem, Feyerabend points out that monistic empiricism rests on a false analysis of experience in its belief that there is a neutral set of theory-independent facts which serve as the ultimate tribunal for settling disputes between scientific theories. Given that one invariably sifts experience through the sieve of some theory or other, it is clear that scientific facts are patently non-neutral. Indeed, there are many 'facts', relevant to the domain of a theory T, which cannot be discovered without the help of an alternative theory T′. Turning the classical view on its head, Feyerabend claims that theories are invented 'in order to criticize observational results'. He rightly points out that many scientists and (alas) some philosophers mistake the predictive utility of a theory for a sign of its truth and reality.

By neglecting alternative theories and the facts they may bring to light, we run the risk of acquiescing in a theory which, though adequate for the phenomena it was devised to explain, is sorely unable to account for anomalies which alternatives would bring to light. For example, this reluctance to take alternatives seriously and to reject all theories still in an embryonic stage enabled Aristotelian physics to survive as long as it did, despite inadequacies which, from the point of view of classical mechanics, should have been glaringly obvious. Feyerabend has dark forebodings of a rigid empiricist orthodoxy in contemporary physics (especially in quantum mechanics) which rejects all theories but the momentarily dominant one because the others are not yet developed in such detail. He argues that a pluralistic science, like a pluralistic society, is the only one which allows for a critical analysis of ideas. The essay is only marred by Feyerabend's habit of exaggerating and caricaturing the view of contemporary empiricists. To say that there are today empiricists who claim that science should be 'built up in a purely inductive fashion' and who refuse to allow hypotheses leaves him open to the charge of exploiting straw-men. Feyerabend does not need to exaggerate the empiricist position to make his critique of it devastating. His remarks on meaning invariance, autonomy of facts and the consistency condition are all valid attacks on current empiricism and it seems pointless to divert his efforts from these fronts simply to beat nineteenth-century empiricism to death.

LAURENS LAUDAN

The Age of Automation. Sir Leon Bagrit. The B.B.C. Reith Lectures, 1964. Weidenfeld and Nicolson, London, 1965. Pp. 86. 15s.

'THE proper function of man is to do that which we do not yet know how to automate.' There have been many and various definitions of the proper function of man, and Sir Leon Bagrit would perhaps not agree with the bald statement just given. It is, however, the conclusion he just avoids drawing in *The Age of Automation.*

Sir Leon sets out to give some characteristics of automation:

(1) It involves feedback: an automated process is designed to give an optimal result.
(2) It extends the range of the senses.
(3) It makes a man the ' real master ' in that he has only the important decisions to make instead of being bogged in a mass of detail.

Thus there is automation in a thermostatic iron (on all counts) but not in a combine harvester. The problems come mostly in connection with point (3). At two separate places in the book these examples are given.

First, the case of an airline pilot flying a supersonic plane. He is assisted by equipment which keeps his plane flying on a satisfactory course and tells him whether there are other planes within risk of collison. He then looks to see whether a collision is likely, and if so, takes avoiding action. He is ' being the fully responsible, fully efficient, human being that automation is allowing him to be '. He is exercising his ' specifically human function '.

Second, the case of a baker's oven man. Apparently it is usual for complicated and expensive baking plants to be operated by a man, whose job it is to keep the temperature and other conditions right. With ' long experience and a sound instinct ' he does quite well, but Sir Leon is clear that a machine could do better. Although the oven man is catering for sudden changes in wind and weather and in the kind of load, etc., he is evidently not exercising any specifically human function.

Clearly the pilot's job is potentially more complex. But coping with the weather in the vicinity of the bakery is not a different *kind* of thing from coping with the weather over the North Atlantic. The only difference seems to be that in the bakery case we can do a completer job of automation. But then look at the pilot. He presumably, when flying his automated plane (if ' flying ' is quite the right word here), takes what decisions have to be taken on grounds of experience, knowledge of flying, and so forth. He is not a random digit. But, if he has relevant experience, then, in so far as he exercises it, he can be automated too. After all, there are algorithms for collision avoidance in two dimensions, and very likely in three.

At this point, of course, there is the temptation to say that the problem is economic: it would be very difficult and far too expensive to make a machine that would perform adequately. However, to say that a specifically human function is one which it is uneconomical to automate sounds almost as bad as the original remark.

It seems clear that the pursuit of a ' specifically human function ' will quickly lead us down the well-trodden road to ' eating, drinking, copulation, evacuation, and snoring ', and quite possibly beyond that to nothing. Perhaps our pursuit is a foolish one, and we should not be led on to it by the reflections which Sir Leon naturally prompts. Instead it may be worth looking at the kinds of activity it is hard to automate.

The salient fact here is that it is hard to automate a man's decisions if his making them is based on particular prior experiences, rather than on aggregates of prior experiences. A computer can readily store more or less gross statistics about past events, and it can in most cases where automation is envisaged find this information without any problems about deciding what to look for. (It should be added that the points are exactly the same if the prior experiences are abstracted into a differential equation rather than a set of statistics.) The situation is quite different when it comes to particular events. In the case where at present automation is insufficient it is because situations arise in the environment which seem to break the rules, i.e. they do not produce clear directives from stored statistics or built in models because the conditions for their applicability do not hold. It is possible for an experienced man to resolve the trouble. It is very likely that what he does is to appeal to his memory of particular prior events, or to memory of his colleagues' descriptions of particular prior events. The resemblance between the present situation and the relevant prior event may be highly tenuous, at least to an outsider—but the vital analogy is present, which means that something like the action taken in the earlier instance may control the present one. Automation in such cases as these is difficult because

(1) We cannot readily store the relevant material. We do not know what it is.

(2) We do not know how to select from the description of an incident that which is worth keeping.

(3) We do not remotely know how to determine mechanically the striking analogy which selects a particular prior event.

These are problems of classification, description, and selection. All these are activities at which people are very good, and towards the automation of which rather little progress has been made. Description is the first necessity, and it is important to note that in the kind of case we are talking about there may never have been any verbal description at all. If I am sailing and a sudden squall arrives from an unexpected direction, I know what to do and do it. The first time such a thing happened I did not know. But it would be a travesty to say that a description had been made of other such incidents and that this had been recorded. The other incidents just happened and were remembered—or partially remembered. Current procedures for classifying, grouping, and selecting do not cater for partial storage: for recording a confused image of a situation as a whole with the addition of some salient aspects which are sharply stored. But as this is written, it occurs that it should not be difficult to devise a procedure which does just that. All that is necessary is to record about an object or situation that it is a typical member of the *n*'th group or class *except* . . . . .

If it were seriously necessary, a classification process could be devised which did this sort of thing and better. In general whatever properties a classifying, grouping, and describing system is required to have, one can propose a model which will at any rate appear to have the required properties. Where then is the catch?

It is twofold. Firstly the models will all be wrong. The person who gave the requirements for the models in the first place will have abstracted from the real situation to say what properties he wants the model to have. He may, as it were, have strained the facts to be able to give a simple statement. The modeller, then, will cook up a system of some sort, and in stating that it has the required attributes he may be a little elastic in his description of it (have there not been computer programs which were described as learning when in fact a stored number was occasionally being increased?) Secondly, even if the model is satisfactory on the first count, it may have an inappropriate internal constitution. I am not saying that its internal constitution must be like that of a man's head: far from it. I am saying that its internal constitution must be such that it can be applied on a realistic scale without giving rise to intolerable computational or constructional problems. It must not depend on exhaustive examination of an exponentially increasing number of possibilities. It may be, but we do not know, that this kind of problem can only be avoided with more knowledge of how human machinery works.

I would suggest that although the pursuit of the limits of automation is profitless as a theoretical exercise, it is useful to look at general characteristics of automation-resistant processes. It is not that these are intrinsically impossible: it is simply that we do not have the rather strong understanding of the processes involved which is needed to make artefacts to perform them.

These notes have concentrated on just one aspect of the book, or rather of the book's subject. *The Age of Automation* discusses the political, social, and educational aspects of automation as well as its nature and purpose. On all of these Sir Leon Bagrit has something to say in addition to the conventional wisdom.

R. M. Needham

Concepts and Principles in the Space-Time Theory within Einstein's Special Theory of Relativity. By H. Törnebohm.
Acta Universitatis, Gothobergensis, 1963. Pp. 98. Sw. Kr. 10.

In this monograph, the space-time of Special Relativity is studied with the aim of illustrating how concepts are introduced into a physical theory and how they are related to measurements. For this purpose, only two dimensional space-time and inertial motion need to be considered.

First, two inertial systems A and B are considered, where B overtakes A. The duration of process OQ as measured in B is shown to have two simple extensions when measured in A: the distance in time from O to Q and the duration of OQ. These provide two explanations of an experiment to verify the clock 'paradox' which motivates postulating P_1: 'The duration of a process is invariant'. P_1, variously expressed, leads to a new definition of simultaneity (σ-simultaneity) which is compared with the ordinary definition (s-simultaneity).

Second, P_2 is introduced: 'The propagation of light is independent of the velocity of the source'. It is shown that P_1 is the simplest assumption about the behaviour of clocks in relative motion such that P_2 is obeyed. In particular, the classical assumption, that distance in time is invariant, contradicts P_2. Distance co-ordinates and hence co-ordinate velocity are defined and Einstein's composition law for velocity follows. With new (null) co-ordinates, the Lorenz transformation equations are found and shown to contain the same information as P_1 and P_2. The invariant interval (distance in space-time) is introduced to complete the signal theory.

Third, another idealised experiment is set up leading to P_3: 'The length of an object measured by rods is equal to its co-ordinate length multiplied by a universal constant', i.e. 'The velocity of light is invariant'. From this, two extensions of the concept of length are introduced, giving two ways of regarding the Einsteinian contraction of a moving body. P_3 is regarded as linking the signal theory of space-time to rigid-body geometry.

The second part of the book constructs space-time deductively. New co-ordinates τ, u, are introduced and the expression for the length of a line segment found in terms of them. The previous relation of τ, u, to experiment is now dropped. One starts with this primitive concept of the length of a line segment, and τ, u, are characterised simply by being in this expression. (This step in the argument could be stated more clearly than it is in the monograph). The principle P is postulated: 'The length of a line segment in space-time is invariant'. Then theoretical definitions are given, in terms of the primitive concept, of the concepts of geodesics, distance in space-time, simultaneity, distance in space and distance in time. Correspondence rules now relate each of these theoretically defined quantities to the experimentally defined concept over certain ranges of the observed quantity. Thus empirical significance is given to the deductive theory in which P_1, P_2, P_3, are now shown to be theorems.

Finally, the role played by the principle of relativity is discussed, and two different forms of the principle are distinguished. These imply postulate P_1 in the construction of the space-time theory. Further, the principle that the fundamental equations of mechanics and electromagnetism are Lorenz invariant connects the descriptive (space-time) theory to the explanatory system of relativity mechanics. It thus makes evidence for the latter evidence for the space-time theory, which links together the theories of relativistic mechanics and electromagnetism.

The concept of σ-simultaneity introduced by the author is defined as follows. Consider a timelike geodesic S and choose an arbitrary point O on it. Then the points of space-time simultaneous with another point P on S are those points P′ such that the distance in space-time δ (O, P′) is the same as that of O and P, δ (O, P). It then follows that we have the same family of spaces of simultaneity defined if we consider any other timelike geodesic S′ through O. However, it seems misleading to call this concept 'absolute simultaneity', as the author does. This is because if we choose any other origin O′ on S for the definition of simultaneity we obtain a different family of spaces of simultaneity, i.e. the choice of a different origin for an 'absolute time co-ordinate' results in a different set of points being called simultaneous with P. In fact, given almost any point P′ spacelike with respect to P, we can find an origin O such that P′ is δ-simultaneous with P.

The author shows in a clear and interesting manner the relation between the different ways of introducing concepts into the theory and of extending operational concepts. Because he has confined himself to inertial bodies, he does not have to consider theories of the type of General Relativity. In providing a basis for Special Relativity, one cannot give *all* the axioms and operational definitions needed for the construction of space-time (which would involve, for example, defining how to recognise light). However, one feels that one further principle and what could perhaps be called a 'statement of intent' might throw some further light on the construction. Thus the author first constructs an experiment showing the clock paradox which he then resolves by defining the structure of space-time *via* the behaviour of light rays. Now one really requires an explanation of why the behaviour of light is of such importance, which it would not be if, for instance, there existed radiation travelling faster than light. Perhaps this could be given by the principle L: 'The speed of light *in vacuo* is finite and no communication is possible at any greater speed'. This then explains the statement of intent I: 'Space-time structure will be investigated by examining the behaviour of light rays.' I is not a principle in the same sense as the other principles, but rather a statement of attitude guiding the argument, that one regards light propagation as fundamental. Such an emendation would, the reviewer feels, bridge a gap of motivation in the author's interesting argument.

G. F. R. Ellis

Methodology of the Behavioral Sciences. By Rollo Handy.
Thomas, Springfield, Ill., 1964. Pp. 182. $6.75.

The methodologies of the behavioural sciences are very diverse, and there is little common ground between, for instance, a Skinnerian behaviourist and a theoretically minded sociologist. It is therefore not surprising that the general principles and recommendations that emerge from this book seem vague or trivial; a useful critical task could nevertheless be performed by contrasting and comparing the various disciplines from the pragmatic point of view adopted by Dr Handy. This is attempted, but the result is disappointing, partly because of a determined adherence to transactionalism, which does little but obscure by giving an air of profundity to mundane points. Thus: ' The advocacy of hard science techniques within a holistic frame of reference, here called transactionalism, is urged because the highly atomistic analytic views often fail to work well. . . . This of course is not to oppose concentration on certain restricted aspects of a transaction, for that may be extremely useful, but rather to point out that such concentration may be mistaken for a full account of the behaviours involved, or lead to neglect of important variables ' (p. 171). But all scientists realise that important variables should not be neglected, and, when they seem not to, can be soundly criticised without introducing transactionalism.

Even without transactionalism the arguments would lack incisiveness, since frequent use is made of sweeping and unsupported generalizations. For instance: ' problems arise because a term having many connotations in ordinary language is specified much more narrowly for scientific purposes. In the previous section, " intelligence ", " rationality ", and " control " are examples. As emphasised earlier, not only laymen, but sometimes the experts then leap to the conclusion that notions previously vague, confused, inconsistent, etc., now have been clarified, and that what was once speculative is now scientific ' (p. 157). No supporting instances are given, though under the heading ' Control ' there is the following: ' Rather rapid shifts from the sense of " control " in which a thermostat regulates room temperature, to the sense in which a dictator controls his populace, are sometimes made. Norbert Weiner, for example, is much concerned over humans becoming victimised by machines, especially since better machines are becoming available ' (p. 153). But this example is neither an instance of a rapid shift in the sense of ' control ', nor of the analogy described in the first sentence of the quotation; and even if it were, it is possible to draw such an analogy without losing sight of its speculative nature.

Among other things Dr Handy minimises the distinction between *conceptual* and *empirical* problems, since both factors may be involved in the discovery of ' warranted assertions ' about human behaviour, and sharp differentiation may lead to misunderstanding; but confusion between them may be even more likely to do this. He emphasises terminological problems, and the way in which terminologies can impede or facilitate scientific inquiry; this could be fascinating, given closely argued analyses of specific examples, but these, as implied above, are lacking. Finally he concludes that there is no theoretical barrier to scientific inquiry into human behaviour; this point is made convincingly, and it is sad that the book as a whole is unlikely to disturb the complacency of the critics of this view.

A. W. STILL

Plato, The Midwife's Apprentice. By I. M. Crombie. Routledge & Kegan Paul, London, 1964. Pp. ix+195. 25s.

HAVING written nearly a thousand pages of closely reasoned exposition and analysis of Plato in the two volumes of *An Examination of Plato's Doctrines* (*EPD*), Crombie now offers a much briefer account in which he presents the conclusions of his study and suggests one interpretation of Plato's significance. The central theme of that interpretation is that Plato conceived philosophy as

> an almost unending struggle to give satisfactory expression to what, in one sense, we know perfectly well all along (p. 75).

But then, in a sense, of course, all philosophy may be said to involve an attempt to make clear what, to begin with, is only dimly understood or not understood at all. Certainly to understand the dialogues we must constantly bear in mind that they represent live conversations, each one a joint endeavour to throw some light on a particular issue. The often circuitous course of those conversations admirably suits Plato's purpose when that consists in examining and clarifying misconceptions as much as in advancing a positive thesis of his own. But if the dialogue form is obviously a source of strength, in that it allows Plato to answer objections and remove possible misunderstandings, on occasions it is also the source of some confusion. And while Crombie rightly commends Plato's conception of the nature of philosophical investigation, he might have been more explicit on the dangers inherent in Plato's style of philosophical writing, which permits, if it does not positively encourage, a certain vacillation and opacity on many important questions, among them the immortality of the soul and the ' population ' of the world of Forms.

Crombie arranges his material in short chapters with such titles as ' the physical world', ' knowledge and belief' and ' the good life '. Once again, as in *EPD*, he writes particularly well in the cause of defending Plato from the charge of ' Platonism ' as that term is popularly understood, and in pointing out where Aristotle's admittedly cursory sketch of Plato's philosophical development needs to be supplemented. In his brief but sympathetic discussion of Plato's attitude towards science (pp. 68 ff.) he rightly argues against the still all too common view that Plato thought that science can entirely dispense with observation. But a rather uncritical attitude towards Plato, which is already detectable in *EPD*, is more in evidence in *PMA*. Crombie's use of inverted commas betrays some special pleading when, having declared that ' it was never part of Plato's message that we should take no interest in the physical world ', he goes on :

> he does indeed demand that we should ' withdraw ' from it, in the sense that we should not be dominated by it ; but this is not in order that we should ignore it, but that we should dominate it (p. 30).

On Plato's political thought, too, Crombie is sometimes quite misleading, as when he says that Plato 'is as little in favour of an inegalitarian society as some of his severest critics' (p. 172), or when he will not allow that Plato could have expected anyone to take the propaganda story (or 'Noble Lie') of the *Republic* seriously (pp. 176 f.). In this book Crombie deliberately omits most of the evidence and arguments on which his judgements (many of them controversial) are based. For those who are content with a summary of his views on Plato, *PMA* presents this in a form that is handy and readable enough. But in general the fuller volumes of *EPD*, in spite of their rather turgid and convoluted style, provide, in my view, both a more stimulating and a better balanced discussion of the topics handled here.

G. E. R. Lloyd

OBITUARY

Reginald Otto Kapp, 2.8.85–20.2.66. It is with deep regret that we announce the death of Professor Reginald Otto Kapp, Chairman, Philosophy of Science Group (later British Society for the Philosophy of Science), 1957–59.

MAN, THE UNIVERSE AND MIND

An Essay in Empirical Analysis

by P. A. MORITZ

"... demonstrates a procedure by which science could progressively take over and systematise such fields of Metaphysics as perception, philosophy of mind, object and causal theory ..."

104 pages. 8/6d.

MARTIGAN PUBLICATIONS

Of all booksellers

MIND

A Quarterly Review of Psychology and Philosophy

edited by Professor **GILBERT RYLE**

with the co-operation of Prof. Sir F. C. Bartlett *and* Prof. C. D. Broad

CONTENTS FOR ISSUE OF JULY 1966

Ten Unpublished Letters from William James, 1842-1910 to Francis Herbert Bradley, 1846-1924	J C. KENNA
To Gödel via Babel	P. J. FITZPATRICK
Aesthetic Perception and the Problem of the "Aesthetic Object"	B. R. TILGHMAN
A Plea for Linguistics	C. G. NEW
"It is True"	JAN SRZEDNICKI
Punishment for Intentions	GERALD DWORKIN and DAVID BLUMENFELD
Weakness of Will	GWYNNETH MATTHEWS
Discussion Notes:	
An Undefined Epistemic Term	HERBERT HEIDELBERGER
Mill and the Naturalistic Fallacy	ALAN RYAN
Causality	R. G. WEYANT
Remarks on Rescher's Note on 'About'	ALAN SLOMSON
A Note on Richard's Paradox	L. J. GOOD
'Suppose Everyone did the same'	H. J. MCCLOSKEY
Critical Notice:	
A Critical History of Western Philosophy. Edited by D. J. O'Conner	JONATHAN BENNETT

NEW BOOKS NOTES

Annual Subscription 25s, post free; single copy 7s 6d, plus postage

Published for MIND ASSOCIATION by

BASIL BLACKWELL 49 BROAD STREET OXFORD

JOURNAL OF THE HISTORY OF IDEAS

A Quarterly devoted to Intellectual History

VOLUME XXVII APRIL–JUNE 1966 NUMBER 2

TABLE OF CONTENTS

ARTICLES

Plato on Writing and Doing Philosophy	*John Fisher*
Descendit Ad Inferos: Medieval Views on Christ's Descent into Hell and the Salvation of the Ancient Just	*Ralph V. Turner*
Saint Augustine on War and Killing: The Problem of the Innocent	*Richard S. Hartigan*
Boyle, Locke, and Reason	*G. A. J. Rogers*
T. H. Green: The Development of English Thought from J. S. Mill to F. H. Bradley	*John Herman Randall, Jr.*
August Bebel and German Social Democracy's Relation to the Christian Churches	*Vernon L. Lidtke*

NOTES

The Aesthetic Principles of the *Peri Hupsous*	*James J. Hill*
The 'Motionless' Motion of Swift's Flying Island	*Robert C. Merton*
The Concept of Denudation in Seventeenth-Century England	*Gordon L. Davies*
A Case of Arms Control in the French Enlightenment	*Seymour L. Chapin*

REVIEW

Walter Kaufmann's *Hegel: Reinterpretation, Texts and Commentary*	*Stephen D. Crites*

Subscription Rates (outside U.S.A. post paid): $6.50 for one year;
£1.75 (plus postage) for single issue

Payable to

JOURNAL OF THE HISTORY OF IDEAS, INC.
The City College, New York, N.Y. 10031

THE REVIEW OF METAPHYSICS

A PHILOSOPHICAL QUARTERLY

Edited by RICHARD J. BERNSTEIN

Volume XIX, No. 4 *Issue 76* *June 1966*

ARTICLES:

Robert S. Brumbaugh	*Applied Metaphysics: Truth and Passing Time*
W. D. Gean	*Reason and Causes*

CRITICAL STUDIES: TWENTIETH CENTURY PHILOSOPHERS

Aron Gurwitsch	*Edmund Husserl's Conception of Phenomenological Psychology*
Richard Schmitt	*Maurice Merleau-Ponty, II*
Howard R. Burkle	*Jean-Paul Sartre: Social Freedom in* Critique de la Raison Dialectique
Peter Achinstein	*Rudolf Carnap, II*

EXPLORATION:

George Armstrong Kelly	*Notes on Hegel's "Lordship and Bondage"*

BOOKS RECEIVED:

Edward A. Reno, Jr. and Staff	*Summaries and Comments*

PHILOSOPHICAL ABSTRACTS ANNOUNCEMENTS INDEX

Annual Subscription $6.00 Student Subscription $4.00

Welch Hall, Yale University, New Haven, Conn. 06520, U.S.A.

Readers resident in Great Britain may subscribe at a special rate through our British representative, Professor Stephan Körner, Department of Philosophy, The University, Bristol 8, England

ZEITSCHRIFT FÜR PHILOSOPHISCHE FORSCHUNG

in Verbindung mit der
"Allgemeinen Gesellschaft für Philosophie in Deutschland"
Herausgegeben von G. Schischkoff, München
Single copy $4; volume (four issues) $13.50

PHILOSOPHISCHER LITERATURANZEIGER

Ein Referateorgan für die Neuerscheinungen der Philosophie und ihrer gesamten Grenzgebiete, in Verbindung mit der "Zeitschrift für philosophische Forschung"
Herausgegeben von G. Schischkoff, München
Single copy $2.30; volume (six issues) $10.80

PHILOSOPHIA NATURALIS

Archiv für Naturphilosophie und die philosophischen Grenzgebiete der exakten Wissenschaften und Wissenschaftsgeschichte
Begründet von Eduard May
Herausgegeben von Joseph Meurers
Single copy $2,75; volume (four issues) $10

VERLAG ANTON HAIN KG., Meisenheim/Glan. (Germany)

Full particulars of the

JOURNALS

published by the

CAMBRIDGE UNIVERSITY PRESS

may be had from

The Manager
Cambridge University Press
Bentley House 200 Euston Road
London, N.W.1

Requests for advertisement rates and data should be made to "Publication Dept." at the above address

THE CONCEPT OF ENERGY

ITS SIGNIFICANCE IN
FUNDAMENTAL THEORY AND PRACTICE

D W THEOBALD

Emphatically not a textbook on energy, this book is rather an attempt to get out bearings on the concept of energy, one of the most important concepts of the physical sciences. The various uses of the word and the development of the concept are investigated, and the possible physical reality of energy is discussed.
This book will be of interest and significance to all working scientists and advanced students for whom the concept of energy is a working tool.

42s

SPON

11 New Fetter Lane London E.C.4

CALCUTTA UNIVERSITY
CENTRAL LIBRARY.

Printed in Great Britain at The University Press, Aberdeen

The British Journal for the Philosophy of Science

CALCUTTA UNIVERSITY
CENTRAL LIBRARY.

Cambridge University Press

August 1966

Volume 17 No. 2 15s. net

Brit. J. Phil. Sci. (1966), *Volume* 17, *Number* 2, 89–104
Printed in Great Britain

CALCUTTA

THEORY AND OBSERVATION (II)*

Marshall Spector

B. The Concept of a Theoretical Term

I shall proceed here as I did in the first part of the paper. First I shall consider analyses of the concept of a theoretical term given by the authors we have been discussing. We shall see that their analyses do not provide the kind of distinction which they require. Then I shall show that the terms 'theoretical' and 'non-theoretical', in some of their common uses, will not provide the type of distinction required either. It will become apparent that these expressions are commonly used in a variety of different ways, to make various points about a term (or object, property, event, etc.). No one of these will be such that all and only those terms and entities which the partial interpretation theorists have *called* 'theoretical' will turn out to be theoretical; and it will not follow that theoretical terms (in each of the senses we shall consider) cannot be directly understood, precluding the giving of semantical rules for them when the theories in which they occur are reconstructed as interpreted logistic systems.

Let us turn, then, to the accounts of the notion of a theoretical term or entity given by the partial interpretation theorists.

1. *The Partial Interpretation Theorists' Analysis*

Braithwaite sometimes seems to mean 'non-observation term' when he speaks of theoretical terms. He writes:

> . . . the initial hypotheses of [a] theory will contain concepts which are not purely logical but which are *not themselves observable* [my emphasis] (call these *theoretical* concepts); examples are electrons, Schroedinger wave functions, genes, ego-ideals.[1]

In other places, it would seem that theoretical terms are to be characterised as those which have no semantical rules; for example, in

---

* Part I appeared in the previous number.

[1] In 'Models in the Empirical Sciences', *Logic, Methodology and the Philosophy of Science*, p. 227

Scientific Explanation (p. 51) he says that the terms ' proton ' and ' electron ' do not have a ' direct meaning '—they are '. . . not understood as having any meaning apart from their place in a . . . calculus '.

Nagel has more to say about what we are to understand by the expression ' theoretical term '; the following passages from *The Structure of Science* (p. 80) appear to sum up what he intends:

> . . . what is intended by this characterization [calling the assumptions about the molecular constitution of liquids a *theory*] is simply that those assumptions employ terms like ' molecule ', which ostensibly *designate nothing observable.* [my emphasis]
>
> On the other hand [in contrast with observation terms], theoretical notions cannot be understood apart from the particular theory which implicitly defines them [*op. cit.* p. 87].

Before commenting on this, let us see what Hempel and others have written on the notion of a theoretical term. In *Fundamentals of Concept Formation* (p. 21), in a section entitled ' The Vocabulary of Science: Technical Terms and Observation Terms ', he writes:

> . . . science has . . . evolved . . . comprehensive systems of *special concepts, referred to by theoretical terms.* [my emphasis]

Later in the same monograph (p. 32), the following passage appears:

> . . . *theoretical constructs,* i.e. the often highly abstract terms used in the advanced stages of scientific theory formation, such as ' mass ', ' mass point ', ' rigid body ', force ', etc., in classical mechanics; ' absolute temperature ', ' pressure ', ' volume ', ' Carnot process ', etc., in classical thermodynamics; and ' electron ', ' proton ', ' Ψ function ', etc., in quantum mechanics.

In a later paper by Hempel, the following explanation of the concepts in question is given:

> . . . *hypothetical,* or *theoretical entities,* i.e., presumptive objects, events, and attributes which *cannot be perceived or otherwise directly observed by us.* [my emphasis][1]

In the glossary of Arthur Pap's *Introduction to the Philosophy of Science,* there is the following entry:

> *Theoretical construct*—concept referring to something that is postulated in order to explain the observed but that is not directly observable (e.g. electron, gravitational field potential, unconscious wish).

---

[1] ' The Theoretician's Dilemma ', *Minnesota Studies in the Philosophy of Science,* **2**, 41

Maxwell seems to maintain that a theoretical entity is simply one which is unobservable; he speaks of entities which *were once* theoretical becoming observable, the implication being that they are *no longer* theoretical [*loc. cit.* p. 14]. Sometimes he seems to suggest that theoretical entities are to be understood as '. . . entities referred to by scientific theories . . .'.

Now almost every one of the passages quoted characterise the notion of a theoretical term in one of two ways: Either a theoretical term is defined as a non-observation term; or, a theoretical term is defined as one which is given no semantical rule in a logistic system reconstruction of the theory in which it is found, and is therefore only understood in the context of the entire theory.

The first of these is just the dual language assumption, which is presently in question. Furthermore, it was shown in Part A that the concept of an *observation term* is itself not as clear as the partial interpretation theorists have believed it to be; there are terms which have been *called* non-observational by the authors in question, and which would therefore be *theoretical* according to the dual language assumption, but which nevertheless *are observational* in several important uses of this expression (e.g. ' is electrically charged '). In brief, we cannot merely *say* that a theoretical term is a non-observation term unless we already have an acceptable characterisation of the notion of an observation term.

The second of these characterisations (i.e. theoretical term = term with no semantical rule) cannot be considered as a *characterisation* (or ' definition ') at all; for we must already know what is *meant* by the expression ' theoretical term ' before we go on to maintain that theoretical terms are to be given no semantical rules in logistic system reconstructions and, moreover, *cannot* be given semantical rules. In other words, when we reconstruct a theory in the manner advocated by the partial interpretation theorists, we must *first* decide in a general manner which terms are theoretical terms, and *subsequently argue* that these terms cannot be given semantical rules in the reconstructed theory.

Nor can we characterise a theoretical entity as one which is referred to by scientific theories. The kinetic theory of gases refers to gases, but a gas is not for this reason a theoretical entity. Nor will it be sufficient to modify this characterisation to ' entities *first* referred to by scientific theories ' (or ' referred to *only* by scientific theories '); for it would have to be the case that scientific theories never introduce terms referring to observable entities, if the dual language assumption is to be maintained. But this is not the case, as can be shown by citing one clear counter example: The term ' plasma ' was *introduced* in a physical theory, and it refers to entities which are observable (in the same way that a lightning bolt is observable).

2. *Carnap on Theoretical Terms*

Finally, let us consider what Carnap has written regarding the notion of a theoretical term. In his recent writings, Carnap does not offer an *analysis* of this notion. A theoretical *term* is one which designates unobservable entities, and a theoretical *entity* is one which is the designatum of a theoretical term.[1] In ' Foundations of Logic and Mathematics ', however, Carnap has introduced some notions which we shall see can be of *some* use in drawing *one* important distinction between various types of entities discussed in physical theories.

Let us begin by noticing that one of the functions of a reconstruction of a physical theory, according to Carnap, is to make clear and precise the meanings of the concepts involved *on the basis of concepts which are already understood apart from the theory being reconstructed.* In particular, one of the purposes of a reconstruction is that of making a theory *intelligible to a layman.* As Carnap puts it:

> For which terms, then, must we give [semantical] rules, for the elementary or the abstract ones? . . . suppose we have in mind the following purpose for our syntactical and semantical description of the system of physics: the description of the system shall *teach the layman to understand it* , . . . [my emphasis][2]

The claim is then made that in order to fulfil this purpose, we *can* give semantical rules for the ' elementary ' (observational) terms of the calculus, but we *cannot* give them for the ' abstract ' (theoretical) terms, since *a layman could not understand them.* (Nor could anyone else, the claim continues; they would be useless transcriptions ' from a symbol in a symbolic calculus to a corresponding word expression in a calculus of words'.[3]) It would appear, then, that we are to distinguish those terms which are understood by what I shall call ' Carnap's layman ', independently of the theory being reconstructed, from those which are not. Generalising the distinction to the ' whole of science '

---

[1] See the quoted passage from Carnap in Part A.

[2] ' Foundations of Logic and Mathematics ', *International Encyclopedia of Unified Science* (University of Chicago Press, 1938), p. 204

[3] Op. cit. p. 210

(as Carnap does), we must distinguish those terms which are understood by Carnap's layman, apart from *any* theory, from those which are not. A theoretical term, then, would be one which is non-elementary in the sense described; and the theoretical-non-theoretical distinction would be identified with the elementary-non-elementary distinction, where the standard observer (or ' standard knower ') is the layman who knows no physics at all.

As it stands, this characterisation has the same defects as Maxwell's characterisation of an *observation* term, taken relative to the ' uninstructed layman ', as discussed in Part A. Either Carnap's layman does not exist, or he must be a savage or an idiot, as Feyerabend has put it. But suppose that we *could* distinguish those terms which (at a given time) are understood by most people (Carnapian laymen) from those which are understood only by physicists (or by scientists in one field or another). The resulting distinction would not be strong enough to support the different treatment given to the two categories of terms so distinguished, by the partial interpretation reconstruction of physical theories. It would be a *mere contingent fact* that there are such laymen, and it would not have been shown that the layman *could not* come to understand ' theoretical terms ' which the physicist understands. After all, physicists were once laymen themselves, the implication being that there is nothing to prevent the interested layman (especially Maxwell's ' sophisticated language user ') from *learning* physics. If this is so, a massive programme of education would leave us with no theoretical terms at all.

It cannot be objected that the two categories of terms distinguished by Carnap's characterisation of a theoretical term *do* in fact differ *conceptually* because theoretical terms cannot be given semantical rules in a logistic system reconstruction of the theory in which they appear. This would amount to *defining* a theoretical term as one which is given no semantical rule in logistic system reconstructions. But I argued earlier that this *cannot* be considered as a *definition* of the expression ' theoretical term ', for it would beg the question at issue. *If* we were to accept the characterization of a theoretical term as simply a term which is not understood by Carnap's layman, the *claim* that such terms cannot be given semantical rules in a logistic system reconstruction would be a *separate claim*, which would need justification. Such a justification has not been given, and I have argued that it *could not* be given merely on the basis of the characterisation in question. The fact that a layman doesn't understand the expressions ' electric field ' and ' atom ' (terms which Carnap would no doubt call ' theoretical ') carries no implication regarding the ' kind of meaning ' such expressions may be given in logistic system reconstructions of the theories in which they occur.

(In the paper of mine referred to earlier, I argued independently that the term ' atom ', as it occurs in the elementary kinetic theory of gases, can—and must—be given a semantical rule in a logistic system reconstruction of the theory.)

We may conclude that Carnap's distinction between elementary and non-elementary terms will not provide the distinction intended by the partial interpretation theorists. (At the very best, he has not shown that the terms so distinguished differ in the manner required, but has merely *said* that they do.) Nevertheless, there is a hint of an important distinction in Carnap's characterisation. In the next section, I shall attempt to point out what this is.

3. *Entities which are not yet understood*

There is an important distinction that can be drawn between objects (properties, relations, events) which are not understood even by scientists, and those which are understood by scientists. Essentially, this amounts to Carnap's distinction relative not to what the uninstructed layman knows, but relative to the state of knowledge at a given time. This is equivalent to the distinction between entities for which we have an acceptable, well-confirmed theory, and entities for which there is as yet no satisfactory theory. For example, the structure of the atomic nucleus is not yet understood by physicists; whereas the structure of the atom is considered to be well understood. This distinction is more readily formulated in terms of non-linguistic entities than it is for terms (or concepts). Physicists do not yet know what the structure of the nucleus is; but it would be misleading to say ' Physicists do not yet understand the concept (or expression) " structure of the nucleus " '. Physicists know generally what is intended by the expression ' structure of the nucleus ' (unlike Carnap's layman, who is supposed not to understand this series of words at all); they do not know *what* this structure is.

The question of the structure of the atom was in this position in the 1920s. We can describe the situation existing at that time in the terminology of interpreted logistic systems by saying that it was not known what the proper semantical rules for terms like ' electron ' and ' psi-function ' should be. In terms of the distinction now under consideration, it could be said that physicists did not understand the nature of the electron. It was a matter of debate as to whether electrons were discrete particles, in which case the psi-function (when written as $\Psi(\mathbf{r}, t)$) would be interpreted as the probability amplitude function for finding such a particle; or whether electrons were extended wave-like entities, in which case the psi-function would be interpreted as representing the charge-density at a point.

It will often be the case—though not always—that when physicists finally do come to an understanding of an object (event, relation, property), it will be through the introduction of new theories, containing new concepts which cannot be completely explained or defined on the basis of the concepts used in previous theories. The solution of the problem of the structure of the atom has been accomplished with the aid of the concept of a discrete quantum of radiant energy, a concept which had no place in pre-twentieth century physics. Thus, conceptual innovation was required to understand the structure of the atom.

But this has not always been the case. An understanding of the internal structure of gases was reached in the nineteenth century on the basis of concepts which were previously understood—the concepts of classical dynamics. Thus the kinetic theory of gases was not so much a ' new ' theory as it was the successful application of an old theory to a postulated domain of objects which satisfied the laws of that theory.

I shall not attempt to solve the problems which arise with respect to the distinction between entities which are not yet understood by scientists and entities which are considered as well understood. I merely wish to point out that it is an important distinction to be drawn, and one which is suggested by Carnap's distinction between terms which are not understood by the *layman* and terms which are so understood. It should be noted, however, that this distinction will not serve the purposes of the partial interpretation theorists. For it will allow as non-theoretical many entities which are called theoretical by the proponents of logistic systems. Only entities discussed at the forefront of research would be theoretical, if this distinction were considered to be identical with the theoretical-non-theoretical distinction of the dual language assumption.

So far I have argued that none of the characterisations of the concept of a theoretical term which have been offered by the authors we have been considering can be used to draw the kind of distinction which they require. These characterisations either (i) do not capture the distinction the partial interpretation theorists seem to have in mind as evidenced by the examples which they give, or (ii) they are not such that it follows that terms which turn out to be theoretical cannot be given semantical rules in logistic system reconstructions of the theories in which they occur, or both. In the remainder of this paper, I shall examine several of the distinctions which are implicit in the common use of the expressions ' theoretical ' and ' non-theoretical '.

4. *Some Common Uses of ' Theoretical ' and ' Non-theoretical '*

I do not believe that a set of necessary and sufficient conditions can be given such that all terms which have been *called* theoretical (by the authors in question as well as by others—including scientists) will satisfy all the conditions, while all terms which are *called* non-theoretical will satisfy none of them. This is the case, I believe, because *there is no one characteristic usually intended by the use of the label ' theoretical '*, and also because *what is intended in any particular instance of the use of this label depends on the context in which it is applied.* The term ' theoretical ' has actually been used to refer to several *distinct* things which may be said of a term or entity, and these characteristics do not, in general, entail one another.

I should like first to list what I believe to be some of the *several* distinctions falling under the general ' theoretical-non-theoretical dichotomy '. Then I shall proceed to discuss them individually. I do not claim that the list is exhaustive. If there are further distinctions that should also be mentioned, this would, indeed, strengthen my thesis. The distinctions which I have in mind are the following:

(*a*) Unobservable—observable.
(*b*) Purely hypothetical—well confirmed as existent.
(*c*) Technical—non-technical.
(*d*) Central to an articulated system—not so central.
(*e*) Unfamiliar to Carnap's layman—familiar to Carnap's layman.

The term ' theoretical ' has actually been used in such a way that sometimes one, sometimes another, of the first member of each of these distinctions seems to be intended. But these distinctions do not all fall neatly together into two classes such that if a term is theoretical in one of these senses, it follows that it must be theoretical in each of the others as well. The ascription of some of these characteristics to a term or entity entails the ascription of some of the others, but they do not all entail each other. We can refer to a particular term or entity as ' being theoretical ' to point out that one of the above characteristics is true of it; and in another context, we can refer to the same term or entity as non-theoretical, thereby making another point about it. There is *some* relation among these various distinctions. Some terms, which we might deem ' clearly ' theoretical, will share several of the characteristics mentioned; and terms which might be recognised as ' clearly ' non-theoretical will share few or none of these characteristics.

It should be emphasised that these are actual uses of the expressions ' theoretical term ' and ' theoretical entity ', and yet the partial interpretation theorists have made no attempt to distinguish them properly.

I shall now explain what I have in mind by each of these distinctions, give some examples, and then draw some general conclusions regarding the notion of a theoretical term or entity.

(*a*) *Unobservable—observable*

This usage of the term ' theoretical ' and ' non-theoretical ' is, of course, the one most frequently intended by the dual language assumption of the partial interpretation theorists. But it was shown in Part A that this distinction itself splinters into several, distinct ones—no one of which will capture the distinction *intended* by the authors we have been considering.

Thus, if an unobservable object is supposed to be an object which can be observed only with high-powered instruments, then viruses would be unobservable, hence theoretical. The satellites of Pluto (if there are any) would also be theoretical in this sense. If ' unobservable-in-principle ' is intended, in the special sense which this expression has in the quantum theory, then all of the elementary particles would be theoretical objects, but there would be a quite precisely defined continuum (of size) from theoretical through non-theoretical objects, beginning with neutrinos and electrons, and going through the other elementary particles, atoms of increasing atomic weight, and molecules of increasing size. Some atoms would be ' less theoretical ' than some molecules. The satellites of Pluto, if there are any, would be non-theoretical objects. If ' observe ' is being used in the sense of ' detect ' (so that the distinction would be ' undetected-detected '), the satellites of Pluto would be theoretical objects. Electrons and neutrinos, on the other hand, would be non-theoretical objects. The reader may provide other such examples.

(*b*) *Purely hypothetical—well confirmed as existent*

This distinction between non-linguistic entities is closely related to the ' undetected-detected ' distinction, but the two are not identical. The point is that good evidence may fall short of actual detection. To illustrate by means of a simple (imaginary) example, suppose that our telescopes were powerful enough to observe satellites for all the planets but Pluto, the latter being too far away. We would then have good reason to believe that Pluto too has satellites, so that they would not be purely hypothetical; yet they would not have been detected. The mountains on the other side of the moon—a few years ago—provides another example of this. There was no doubt whatsoever that there were mountains on the other side of the moon, but they had not actually been detected. The neutrino provides an interesting borderline case. Some physicists considered neutrinos as *not* being purely hypothetical even before the experiments in which it is said that they were finally detected. Others were not convinced until these experiments were successfully carried out. The case of some double stars provides a rather interesting example of the same point. If an astronomer finds a star which shows a certain type of change of luminosity with time, or which seems to wobble in a periodic manner, he will say that he has detected a double star; yet he may also say that he has only detected *one* of the stars of the pair. The other star, however, will *not* be considered purely hypothetical.

The distinction in question is related to the distinction discussed in Section 3 between objects (events, relations, properties) which scientists do not yet understand, and objects which are considered as well understood. For it will often be the case that objects posited to explain other objects (events, etc.) which are not understood will be, for a time, purely hypothetical. The neutrino, which was posited to explain beta-decay phenomena, is an example of this.

Only (but not all) entities posited by the most advanced stages of physical theory would be theoretical in this sense.[1] This is the use of ' theoretical ' in which we say ' *only* a theory as yet '. Clearly, this distinction will not serve the purposes of the authors in question. Almost all of their examples of theoretical entities would have to be non-theoretical in the present sense. Also, the fact that an entity is considered to be purely hypothetical does not imply that we cannot give semantical rules for terms referring to the entity and its properties. (Indeed, some purely hypothetical entities may have quite familiar *sorts* of properties—e.g. men on Mars.)

---

[1] An example would be the ' geon '; cf. John Wheeler, ' Curved Empty Space-Time as the Building Material of the Physical World: An Assessment ', in *Logic, Methodology and the Philosophy of Science*, 368

(*c*) *Technical—non-technical*

A technical *term* is one which has a special use in a particular discipline. Such terms will often be precisely defined for certain purposes—usually of a systematic nature. As the term 'speed' is usually used in ordinary circumstances, it is a non-technical (non-theoretical) term. If I say 'The speed limit here is thirty miles per hour', I have not employed any technical terms. On the other hand, the expression 'instantaneous velocity' is a technical term from classical mechanics, where it is precisely defined as a vector quantity which is the limit of a certain sequence, which is in turn precisely defined on the basis of other expressions.

There are many terms which are used both in ordinary (non-technical) contexts, and in technical contexts. Some of these will be terms found in common use, which have been taken over by scientists for certain systematic purposes. Others will have been first introduced by scientists in technical contexts, and subsequently will come to be used in ordinary, non-technical, situations. For terms such as these, we can speak of technical and non-technical *uses*. The term 'insect' provides an instructive example of a term of this sort. It will be worthwhile to quote from a dictionary. (Notice that the distinction is drawn in such a way that the technical use of 'insect' is considered to be the *correct* ('strict') use. This need not be done; it indicates, I think, a suggestion that any well-educated person is, or should be, familiar with the technical use of the term.)

> in'sect . . . 1. Strictly (*Zool.*), any of a class (Insecta) of small invertebrate animals, with three clearly defined body regions, *head*, *thorax*, and *abdomen*, with only three pairs of legs, and usually with wings, as beetles, bugs, flies, etc. 2. Popularly, also, any of certain allied classes of arthropods whose members are wingless and usually have more than six legs, as spiders, mites, ticks, etc. . . .[1]

---

[1] *Webster's New Collegiate Dictionary* (Springfield, Mass.: G. & C. Merriam Co., 1959), p. 434

The second entry appears to be as technical as the first, but what it does is to delineate in zoological terms the class of objects which are often (' popularly ') referred to as insects by those who are not familiar with the technical use of the term ' insect '.

Another example of a term of this sort would be the term ' weight '. It could be used in its technical or non-technical sense depending on whether we are referring to a bathroom scale reading (for example) or to the mass of a body multiplied by the acceleration due to gravity upon the body. (The two are related, of course; otherwise physicists would have used a different expression for their purposes.)

In *Fundamentals of Concept Formation* (p. 32), Hempel uses the term ' volume ' as an example of a theoretical term. In terms of the present distinction, it could be said that he has in mind the technical *uses* of this term in physics, as a quantity precisely defined in terms of the linear dimensions of a body.

It should be clear that the present distinction will not serve the purposes of the partial interpretation theorists. To use Hempel's example, is it true that the term ' volume ' can only be understood in the entire context of the theory (theories) in which it occurs? (Are ' volumes ' unobservable?) Something has gone wrong if we must draw conclusions such as this, based on the dual language assumption.[1]

(*d*) *Central to an articulated system—not so central*

An expression is central to a system if changes in its definition would force an important modification of the system, or, if it enters into many or ' key ' relations in the system. Such an expression would be said to be of ' great theoretical significance '. This use of the term ' theoretical ' is closely related to the one just discussed, but they are not identical. For a concept may be quite precisely delineated and yet be of no systematic significance. Thus the expression ' kinetic energy of a body ' (defined as one-half the mass of the body multiplied by the square of its velocity) has great ' theoretical significance ' (because it occurs in conservation laws), while the expression ' cubic momentum of a body ' (defined as the velocity of a body multiplied by the cube of its mass) has none. Both of these expressions are quite precisely defined; yet the former is central to an articulated system (classical dynamics), while the latter is not. The term ' electron ', at the end of the nineteenth century, was not a theoretical term in this sense; it has since become one. If it had been found in 1900 that there were several types of electrons, with different masses (say), such a discovery would not have affected physical theory in any serious manner; but if the same discovery were made today, quantum theory would have to be modified quite drastically.

---

[1] The reader who has caught the drift of my argument to this point may be tempted to add the following argument: '*All* terms in a semantical system reconstruction of a theory would turn out to be theoretical. We can speak of the theoretical counterparts of ordinary expressions in such reconstructions. For example, if a term " G " in such a system is given the semantical rule " ' G ' designates the property of being green ", " G " would be a theoretical (technical) term, but " green " would not.'

But this would be unfair, for we are to suppose that we are examining what I have called ' unreconstructed theories ' in order to pick out those terms which are theoretical and those which are not. It is after we have done this that we give the logistic system reconstructions, in which all of the terms will be precisely delineated.

I do not claim to have explicated in any precise manner what is involved in the notion of being central to an articulated system. I merely wish to point out that the distinction in question is an important one, which has been made, and has been referred to in terms of the ' theoretical significance ' of an expression. It is clear, however, that the distinction in question will not serve the purposes of the authors we have been considering. I can define the term ' selectron ' as an elementary particle having the mass of the electron, twice its charge, spin equal to zero, etc.; such a term would have no theoretical significance whatsoever (because there is nothing which has these properties), but the partial interpretation theorists would presumably call this a theoretical term. (If they would refuse to call this a theoretical term because there are no such things, they would also have to maintain that we *did not know* whether the term ' neutrino ' was a theoretical term before the experiments which established their existence.)

(*e*) *Unfamiliar to Carnap's layman—familiar to Carnap's layman*

In this rather uninteresting use of the term ' theoretical ', any expression introduced in any context whatsoever which has not yet come to be used universally for some purpose would be theoretical. It is an accident of our educational system, as it were, that there are any theoretical terms at all.

The most that can be said about this use of the term ' theoretical ' is that it provides a (very) rough first approximation, or introduction, to some of the problems in the analysis of science. Thus we might say that one of the problems of the philosophy of science is: Why do scientists often introduce terms which the ' layman ' has never heard of, and has no occasion to use in ordinary circumstances—' theoretical ' terms?

These, then, are some of the several points that can be, and have been made by referring to a term or entity as theoretical or non-theoretical. Although the resulting distinction are quite different distinctions, they are related in several way. Historically, entities which cannot be observed without the aid of high-powered instruments were at one time purely hypothetical. But this need not always be the case; a series of experiments may be considered as affording overwhelming evidence for the existence of some new entity, even at the time of the experiments and before the development of instruments which can be used to observe the entity in question. That is, there need not be a period during which the unobservable entity which explains the results of the experiments is considered to be purely hypothetical. I have already pointed out that undetected entities are usually—but not always—purely hypothetical. Highly technical terms will usually be unfamiliar to Carnap's layman.

Terms or entities which in some intuitive sense are recognised as particularly clear cases of being theoretical will often share all or almost all of the characteristics discussed. Many of the examples of theoretical terms used by the partial interpretation theorists are of this type. (I suspect that this is the source of any plausibility which their view may have.) All of the authors in question cite the elementary particles of modern physics as (paradigm) examples of theoretical entities. The neutrino at the time of the introduction of the neutrino hypothesis provides a particularly striking example of this. The neutrino was posited to explain a *prima facie* violation of the principle of the conservation of energy in beta-decay phenomena. It was thus invoked to explain a type of event which was not then understood by physicists. At that time, neutrinos were purely hypothetical, unobservable-in-principle (in the special sense discussed in Part A) undetected, and quite unfamiliar to Carnap's layman. The term ' neutrino ' was also a highly technical expression, being quite precisely defined. Neutrinos have since been detected, and are no longer considered as being purely hypothetical (the latter may have taken place first). Since we still lack a comprehensive theory of elementary particles, they are still not considered as being well-understood by physicists. And, of course, they are still unobservable-in-principle, unfamiliar to Carnap's layman, and the term ' neutrino ' is still a technical one—even more so.

At the other extreme there are terms or entities which might be recognised as particularly clear cases of non-theoretical terms or entities. Such terms or entities will share almost none of the characteristics discussed; e.g. tables (or ' table '). And then there are terms or entities for which there might be some disagreement as to whether they should be classified as being theoretical or non-theoretical. The term ' virus ' (or viruses) might be an example of such a term. But arguments as to whether viruses are ' really theoretical ' or not lose their force when it is recognised that some of the characteristics mentioned are true of them, while other are not. That is, in one context we may refer to viruses as theoretical entities in order to make one point about them, and in a different context, we may call them non-theoretical entities to make a different point about them. At one time, they could not be observed with any instrument, but they have since become observable with the aid of the electron microscope. Hence they have been detected. Their existence has become quite highly confirmed, whereas they were once purely hypothetical. The term ' virus ' was once not central to any well-articulated theory, but it is rapidly becoming so with advances in micro-biology. (In this sense, the term ' virus ' is becoming more theoretical.) The term ' virus ' is now becoming familiar, to some extent, to ' Carnap's layman '. Finally, with advances in micro-biology and organic chemistry, the concept of a virus is becoming more precisely delineated, where formerly a virus was thought of as a very small (non-filterable) organism; thus the term ' virus ' is coming to have quite technical uses.

The point is that once these facts about viruses (or any other type of entity) have been spelled out, argument as to whether or not ' virus ' is a theoretical term is *at best* argument as to which of the features described is ' most impressive '. And even this will depend on context and purposes. There is no one property which is the ' essence of theoreticalness ', and which will allow us to decide in a yes-no-manner, apart from any context or purpose, whether or not ' virus ' ' really ' is a theoretical term. Once these characteristics have been parcelled out for a given term or entity (and this can be done without even using the term ' theoretical '), there is no more to be done. ' But is " virus " theoretical or not?' ' Suit yourself! '

5. *Conclusions*

The conclusions of Part B may be summarised in the following way: First, I have pointed out that the partial interpretation theorists have not given us a clear and precise characterisation of what they have in mind when they refer to a particular term or entity as ' theoretical '.

The most they have done is either (*a*) identify the expression 'theoretical' with 'unobservable' (or 'non-observational'); or (*b*) say that a theoretical term is one which is given no semantical rule in a logistic system reconstruction of the theory in which it occurs, thus making it understandable only in the context of the entire theory. The first we saw to be unclear, because they have not characterised the notion of an observation term in an adequate way. The second, I argued, cannot be considered to be a *characterisation* in the first place. *At best*, it *assumes* what must be *established*; at worst, it is not even true that this is an assumption, because we do not know what a theoretical term *is*, except a term which is to be treated as described.

Secondly, I have argued that the term 'theoretical' has in fact been used (by scientists and philosophers) to make several distinct points about a term or entity. To paraphrase the conclusion of Part A; if we hear the sentence 'Xs are theoretical' out of context, we still do not know what exactly is being said about Xs. We would have to know who has said it to whom for what purpose; and depending on these latter factors, the assertion may be true or false, even for one and the same entity. In other words, there is no such thing as a (context and purpose independent) *theoretical term* or *theoretical vocabulary*.

If the considerations which I have offered carry any weight, we may conclude that the two basic assumptions of the partial interpretation version of the interpreted logistic system approach to the analysis of physical theories—the dual language assumption and the assumption that 'theoretical terms' (these are now raised-eyebrows quotes) cannot be directly understood via semantical rules—are incorrect.

Duke University,
Durham, No. Car., U.S.A.

Brit. J. Phil. Sci. (1966), *Volume* 17, *Number* 2, 105–126
Printed in Great Britain

A LOGIC FOR EVIDENTIAL SUPPORT (II)*

L. JONATHAN COHEN

VII

IN the first part of this paper (this *Journal*, 1966, **17** (65), 21 certain informal conditions were postulated to which a confirmation-functor denoting degree of evidential support by eliminative induction should conform. Various consequences were shown informally to follow from these conditions, including a law which excludes such a functor from being an interpretation of any well-formed expression within the classical calculus of probabilities. It was also argued that these conditions and their consequences were implicit in at least one familiar concept of evidential support and could be illustrated from its actual operation in science.

It is possible to justify informally a number of other syntactical conditions to which confirmation-functors for eliminative induction must conform. But their standing would remain relatively obscure and questionable if it were not possible to axiomatise their syntax so that the various separately established laws achieve a certain amount of mutual corroboration and their exact implications may be fully appreciated. Moreover it would be best if this axiomatisation were achieved within a fairly familiar type of formal system. Probabilistic theories of evidential support have at least the merit of connecting the relatively unfamiliar and controversial concept of a confirmation-functor with a calculus that articulates the structure of many other, well-accepted bodies of thought as well. Something similar is desirable for any non-probabilistic theory of evidential support, and in fact this can be provided by a certain kind of generalised modal logic. So far as their syntax is concerned confirmation-functors will turn out to be reducible to, i.e. definable as abbreviations of, expressions containing nothing but the symbolism of such a logic.

I shall now explain certain features of this system, before setting it out in detail.

It is already known that C. I. Lewis's systems of strict implication may be generalised.[1] In this way it is possible to represent not only Lewis's standard modality 'it is logically necessary that', but also any desired number of other modalities of certain kinds, such as 'it is analytically necessary that' or 'it is physically necessary that'. It is also possible, within such a system, to distinguish any desired number of different kinds of non-extensionality. Contexts that resist the substitution of truth-functional equivalents, for example, may not be equally resistant to the substitution of logical or analytic equivalents. How is such a system achieved? Instead of including just a single type of modal-operator in our primitive notation, as in Lewis's system, we include any desired number of modal operators, each with its differentiating numerical superscript, viz. $\Box^1$, $\Box^2$, . . . $\Box^n$. Lewis's axioms are then generalised correspondingly. For example, in the calculus O4 we have $\Box^n A \rightarrow^d \Box^n \Box^n A$ as an axiom-schema, where $B \rightarrow^d E$ is an abbreviation for $\Box^d -(B \,\&\, -E)$ and d is some designated positive integer. We also have $\Box^n A \rightarrow^d \Box^m A$ where $d>n=(m+1)$, so that we arrange all the square-operators we need for any particular interpretation in a hierarchy of subalternation, with $\Box^d$ at the top.

---

* Part I appeared in the previous number.

But what is the lower limit of such a hierarchy? In O4 any wff A was allowed to be written $\Box^\circ A$, so that the axiom-schema $\Box^n A \rightarrow^d \Box^m A$ included within itself the axiom-schema $\Box^n A \rightarrow^d A$. In effect this treats 'it is true that' as the null case of modal prefix. But we certainly do not need to cut off the descending order of modal operators at this point. Just as 'A is logically necessary' or 'A is analytic', or 'A is a law of nature', asserts that A is of stronger standing than would be conveyed by the mere assertion of A or of 'A is true', so too there are expressions capable of asserting that A may be of weaker standing than this. If we say, for instance, 'There exists at least two-thirds of the reasons that would suffice to justify anyone's accepting the truth of A', we leave it open whether A is actually true or not. Thus below the lowest modal operator—$\Box^1$, perhaps—which is such that the statement made by prefixing it to A logically implies A, we may still place any number of other operators. At the lower limit of this hierarchy we could now place an operator $\Box^\circ$, such that $\Box^\circ A$ is no longer a suitable way of rewriting A but is available for a much weaker interpretation, e.g. 'There exist at least zero of the reasons that would suffice to justify anyone's accepting the truth of A'. But A itself will no longer have any position in the hierarchy. We shall simply have two designated superscripts, d and c, where $d \geqslant c > 0$, such that $\Box^{d}A$ is the strongest assertion of A, $\Box^{c}A$ is the weakest assertion of A that still implies, without being implied by, the statement that A is true, and $\Box^{0}A$ is just the vanishing-point of A-assertion. In other words, 0 denotes the level where there is nothing to be said on behalf of a statement, c denotes the level where there is sufficient to be said for it to be regarded as having attained the threshold of reliability, and d denotes the level of logical truth.

---

[1] L. Jonathan Cohen, 'A Formalisation of Referentially Opaque Contexts'; *Journal of Symbolic Logic*, xxv (1960), pp. 193 ff.

It is convenient to take $c = 1$, since this allows us to attach to $\Box^{n}A$, where $c > n$, the meaning 'There is at least a fraction n of the reasons that would suffice to justify accepting A'. We shall turn out thereby to be able to retain the usual convention of confirmation-theory, that if $c[H, E] = 1$ then H is true. But it is by no means indispensable to put $c = 1$. We might, e.g. simply *count* reasons rather than *measure* evidence, and interpret $\Box^{n}A$, where $c > n$, as 'There are on balance at least n reasons for accepting A'. Or we might put $c = 100$, and interpret $\Box^{n}$ in terms of percentages, where $c > n$.

It might be thought that instead of assigning the intermediate position to $\Box^{c}A$, because $\Box^{c}A$ still implies the truth of A, we should economise by allowing A to be rewritten as $\Box^{c}A$ and so assigning the intermediate position to A itself. The axiom-schema $\Box^{n}A \rightarrow^{d} \Box^{m}A$, where $d \geqslant n > m$, would then still include within itself $\Box^{n}A^{d} \rightarrow A$, where $n > c$. But we should then also be able to prove $A \rightarrow \Box^{n}A$, where $c > n$, and it is not clear that the mere truth of A implies the existence of any reasons at all for accepting A. Admittedly 'A is logically true' or 'A is analytic' or 'A is a law of nature', in an appropriate sense, implies, e.g. 'There exist on balance at least two thirds of the reasons that would justify anyone's accepting the truth of A'. But does just 'A is true' imply this? We shall never want to claim that A is logically true, or analytic, or a law of nature, without supposing there to be factors not identical with A that are reasons for accepting A, e.g. intuitive notions of valid argument with which A conforms, or meanings that make A's truth necessary, or experiments that demonstrate A's operation, respectively. But we may conceivably want to claim that A is true without thereby committing ourselves to the existence of reasons for accepting A. For example, we may want to make this claim because we think that belief in the truth of A should be a matter of faith rather than reason, or—more relevantly to the problem of eliminative induction—because the only kinds of positive reasons in which we are ultimately interested are observed coinstantiations, in certain circumstances, of the antecedent and consequent of a generalised conditional statement, while one of the hypotheses in which we are interested is true just because its antecedent is never in fact instantiated. Accordingly, if we are to extend our hierarchy of modal operators to lower levels, at which $\Box^n A$ no longer implies A, we must regard A itself as being quite outside the hierarchy, so that we nowhere have it as a law that A implies $\Box^n A$. We shall need thus to be able to prove both $\Box^n A \rightarrow \Box^m A$, where $d > n > m$, and also $\Box^c A \rightarrow A$, where $d > c$. And we should also do well to make available a non-denumerable infinity of square-operators as primitive terms, in order not to inhibit unnecessarily the statement of just how much of the requisite evidence there is for accepting A. (The exact form a confirmation-metric should take is a matter for empirical science, not formal logic.)

Certain other alterations will follow on from this downward extension of the hierarchy of square-operators.

First, if A is logically true we shall want it to be a law that there are no reasons for accepting $-A$. That is, if A is provable in our system we shall also want to be able to prove $-\Box^n - A$. Now, where $n > c$ we have $\Box^n A \rightarrow A$, and it is then easy enough to show (by contraposition and detachment) that if we can prove A we can prove $-\Box^n - A$. But since we do not have $\Box^n A \rightarrow A$ where $c > n$, we cannot then show that if A is provable so is $-\Box^n - A$. We need therefore to make special provision in our axioms to ensure that if A is provable so too is $-\Box^n - A$ where $c > n$.

Secondly, it will not be useful to be able to prove $(A \rightarrow^n B) \rightarrow (\Box^n A \rightarrow^n \Box^n B)$—as is provable for all n in O4—where we have $c > n$. We shall want rather to be able to prove $(A \rightarrow^n B) \rightarrow (\Box^n A \rightarrow \Box^n B)$ there, as is demonstrable from the Consequence Condition (2) and the General Conjunction Condition (10). Let us suppose that E is a complete description of the universe. Then we have, on our proposed interpretation for $\Box^n$ where $c > n$,

(13) $((A \rightarrow^n B) \& \Box^n A) \rightarrow c[A \rightarrow B, E] > n \& c[A, E] > n$

Case i: assume $c[A \rightarrow B, E] = n = c[A, E]$

Then from (13), (10) and (2) we have

(14) $((A \rightarrow^n B) \& \Box^n A) \rightarrow \Box^n B$

Case ii: assume either $c[A \rightarrow B, E] > n = c[A, E]$ or $c[A \rightarrow B, E] = n < c[A, E]$

Then again from (13), (10) and (2) we have (14).

Case iii: assume $c[A \rightarrow B, E] > n < c[A, E]$. Then there is some m, such that $m > n$, and we have either $c[A \rightarrow B, E] = m = c[A, E]$ or $c[A \rightarrow B, E] > m = c[A, E]$ or $c[A \rightarrow B, E] = m < c[A, E]$. In each of these three cases (13) and (10) will yield us $((A \rightarrow^n B)\ \&\ \Box^n A) \rightarrow c[(A \rightarrow B)\ \&\ A, E] = m$ and thence with the aid of (2) we may derive (14).

Finally, by exportation, (14) gives us $(A \rightarrow^n B) \rightarrow (\Box^n A \rightarrow \Box^n B)$.

Thirdly, what happens to $\Box^0 A$ if it is no longer introduced as a method of rewriting A? It should now be available for such an interpretation as: 'There exists at least zero of the reasons that would suffice to justify accepting A.' But since at least zero evidence will exist for A, whatever A is, $\Box^0 A$ should now be demonstrable in our formal system. In fact a suitable wff is available for abbreviation as $\Box^0 A$, once we introduce quantification over our modal operators. Let us postulate that all the square-operators occurring as primitive symbols have superscripts denoting real numbers greater than zero. Then an expression which is to have the meaning that *at least* zero evidence exists for A may be treated as an abbreviation for the assertion that either there is some evidence for A or there is none. That is, where $\Box^x$ is a variable taking primitive square-operators as its substitution-instances, $\Box^0 A$ is most appropriately treated as an abbreviation for the tautology $\exists\Box^x(\Box^x A) \vee -\exists\Box^x(\Box^x A)$.

VIII

The calculus O4 was a generalisation of the Lewis-Barcan calculus S4. When the hierarchy of square-operators is extended downwards, with consequential modification of the axioms and abbreviations, and quantification is introduced, with appropriate axioms, over square-operators, as well as over individual letters, let us call the resultant calculus $O4^4$. I shall now propose certain definitions of confirmation-functors in terms of the symbolism of $O4^4$. I shall then explore the syntax of such confirmation-functors by developing $O4^4$ in detail. It will emerge that these functors conform, in principle, to all the conditions that were discussed informally in I to VI above.

$\Box^n A$, as we have said, may represent the statement that there exists at least n of the reasons sufficient to justify accepting A. But of course this is merely a statement about what exists on balance in the universe as a whole, whereas actual statements of evidence report what has already been observed at known places and times. So what confirmation-functors are useful for is to tell us which statements of the $\square^n A$ kind we may infer from certain kinds of evidential statement. As a first approximation, therefore, we might consider defining $c[H, E] \geqslant n$, at least where $n > 0$, as $E \rightarrow \square^n H$. But this would be unsatisfactory because then, for any E and any H, if E were false $c[H, E] \geqslant n$ would be true for any choice of n whatever: i.e. a hypothesis could be said to derive an unlimited degree of support from a statement that was in fact quite irrelevant to it. In order to avoid thus getting entangled here in the paradoxes of material implication we need to define $c[H, E] \geqslant n$, where $n > 0$, not as $E \rightarrow \square^n H$ but rather as $E \rightarrow^c \square^n H$. We can then define $c[H, E] > 0$ as $\exists \square^x (E \rightarrow^c \square^x H)$, $c[H, E] = 0$ as $-(c[H, E] > 0)$, and $c[H, E] \geqslant 0$ as $c[H, E] > 0 \vee c[H, E] = 0$. Similarly $c[H, E] \leqslant n$, where $n > 0$, may be written as $\forall \square^x ((E \rightarrow^c \square^x H) \rightarrow (\square^n H \rightarrow^d \square^x H))$, and $c[H, E] = n$ as $(c[H, E] \geqslant n) \& (c[H, E] \leqslant n)$. It will also be possible in $O4^4$ to produce definitions for merely comparative confirmation-ratings. We can have $c[H, E] \geqslant c[I, F]$ as an abbreviation for $\forall \square^x ((F \rightarrow^c \square^x I) \rightarrow (E \rightarrow^c \square^x H))$, $c[H, E] > c[I, F]$ for $-(c[I, F] \geqslant c[H, E])$, and $c[H, E] = c[I, F]$ for $(c[H, E] \geqslant c[I, F]) \& (c[I, F] \geqslant c[H, E])$.

We do not want our calculus to require $c = d$, because we want to keep open the possibility of regarding our confirmation-metric as being subject to revision in the light of experience. At the same time we must define $c[H, E] \geqslant n$ as $E \rightarrow^c \square^n H$, and not as $E \rightarrow^m \square^n H$ where $c > m$, since to use a confirmation-metric at all we must assume it has sufficient support to be reliable (even if on another occasion we may want to be able to assess the actual extent of this support). So far as $\square^n H$ is concerned, however, there need be no syntactical restriction on n in relation to c. A confirmation-metric might conceivably assign to a hypothesis, on the basis of certain evidence, not only a lower but also a higher degree of support than suffices for it to attain the prescribed threshold of reliability. As evidential support by eliminative induction accumulates beyond the point at which it suffices to establish our hypothesis as a law, we may conclude informally therefrom that the law holds over a wider and wider domain of application, and we are thus entitled to assume that it is getting safer and safer to entrench the law-sentence in our language as an analytic statement. Hence even in relatively advanced sciences we may still need a syntax of evidential support in order to deal with measurements of a theory's range of application, for this kind of measurement is continuous in principle with ordinary assessment of evidential support.[1] We may therefore have $c[H, E] > n$ where $c \gtreqless n$.

Thus instead of the normal probabilistic thesis that $c[H, E] = n$ puts the relationship of H to E at a certain point between deducibility and contradiction, I am proposing that such a confirmation-rating should be taken to put the standing of H, if E is true, at a certain point between logically necessary truth and utter worthlessness. Indeed the logical syntax of evidential support is here modelled throughout, *mutatis mutandis*, on that of logical necessity. In particular we may note now a feature of $O4^4$ that involves a controversial issue in modal quantification theory. In non-modal quantification theory we have $A \rightarrow \forall x(A)$ as an axiom or theorem only where x is not free in A—for the obvious reason that, where x is free in A and is regarded as capable of functioning as an individual constant, $A \rightarrow \forall x(A)$ would appear to authorise generalisation from a single instance. It would appear to authorise inferring 'Everyone is wise' from 'Socrates is wise'. If however A is an axiom or theorem of our logic, we are rightly allowed in non-modal quantification-theory to derive the closure of A, since truths of logic are supposed to remain true under all uniform replacements of their non-logical terms. Hence we may regard it as a truth of logic that if A is a truth of logic $\forall x(A)$ is true: if it is a truth of logic that if Socrates is a Greek philosopher Socrates is a Greek, then any Greek philosopher is a Greek. It follows that, if $\Box A$ is to represent 'A is a truth of logic', and all truths of logic involving quantifiers are to be represented by axioms or theorems in modal quantification theory, then $\Box A \rightarrow \forall x(A)$ ought to be provable in modal quantification theory, even where x is free in A. Consequently, where we can also prove $\Box(\Box A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$, as in S4, we ought to be able to prove $\Box A \rightarrow \Box \forall x(A)$: it is not sufficient just to have a rule of derivation entitling us to infer that if $\Box A$ is provable so is $\forall x(\Box A)$. Moreover, if we can prove $\Box A \rightarrow \Box \forall x(A)$, we shall be able to derive from it, by standard rules of quantification, what Prior called the Barcan formula,[2] viz. $\forall x(\Box A) \rightarrow \Box \forall x(A)$. Hence modal systems in which the latter is not derivable[3] seem to be inadequate as formalisations of the most important modality in logic, the modality of logical truth itself.

---

[1] Cf. L. Jonathan Cohen, *The Diversity of Meaning* (1962), pp. 311 ff.

[2] A. N. Prior, *Time and Modality* (1957), p. 26.

[3] E.g. S. A. Kripke, 'Semantical Considerations on Modal Logic', *Acta Philosphica Fennica* xvi (1963), pp. 83 ff. I am not denying that there are some interpretations or models for S4 with which the Barcan formula conflicts. Obviously the intuitionistic interpretation is one such. But in the interpretation proposed for it by its

Correspondingly, if $O4^4$ is to articulate a logical syntax of evidential support that models itself, *mutatis mutandis*, on the syntax of logical necessity, $\Box^n A \rightarrow \forall x(A)$ ought to be provable where $n \geqslant c$, but not where $c > n$, since this would entitle us to infer the truth of A from even the very slightest amount of favourable evidence for A.

Of course we must be careful to place appropriate restriction on the interpretation of any modal calculus that contains such theses. For example, if free occurrence of some individual letter x may be taken to represent 'Socrates', and free occurrences of some predicate letter S to represent 'is identical with Socrates', then, since $\Box Sx$ is true and $\Box Sx \rightarrow \Box \forall x(Sx)$ is provable, we shall be able to detach 'it is logically true that everything is identical with Socrates'. There are various ways to avoid legitimising such deductions of false conclusions from true premisses. We might prohibit any individual letter of the calculus from being interpreted both as a constant, when free, and as a variable, when bound. Then either we should include in our primitive notation two lists of individual letters, one lot to function as individual constants and the other as individual variables (bound or free), along with appropriate axiom-schemata. Or, more simply, all the individual letters would function as variables, bound or free, and when we wanted a formula of the calculus to represent a statement in which a proper name, personal pronoun or demonstrative was used referentially, we should use an appropriate definite description to represent this referring expression instead of an individual constant. But neither of these two procedures would enable $\Box A \rightarrow \Box \forall x(A)$ to represent 'If it is a truth of logic that if Socrates is a Greek philosopher Socrates is a Greek, then it is a truth of logic that any Greek philosopher is a Greek'. To achieve this we need to be able to interpret individual letters both as constants, when free, and as variables, when bound, as we are able to do for non-modal quantification theory. We can achieve such an interpretation for $O4^4$ if we take free occurrences of individual letters to represent designations of individuals in terms that are completely absent from, and analytically unrelated to, the predicative expressions that we take our predicate letters to represent.

originator, C. I. Lewis, as a theory of logical possibility and logical necessity (cf. C. I. Lewis and C. H. Langford, *Symbolic Logic*, 1932, p. 153), the Barcan formula is essential to S4, so long as a statement is to be held logically true if and only if it remains true under all uniform replacements of its non-logical terms.

For example, if individual letters, when unbound and functioning as constants, represent proper names, then no proper name may be part of what is represented by a predicate letter and no statement containing a proper name may be analytically true unless it remains analytically true under all uniform replacements of that name. If instead the elements of our universe of discourse are identified by spatio-temporal references, then the terminology of such references must be similarly unrelated to what is represented by any predicate letter. Nor should either restriction be any hardship in a calculus of confirmation where we are primarily concerned with evidential support for the universal hypotheses of natural science. Normally, as Popper has emphasised,[1] no proper name, or reference to a particular individual, occurs in such a hypothesis. Hence no individual reference need be made by any of the predicate letters which, along with quantifiers, bound variables, and logical connectives, compose the symbolic representations of these hypotheses.

It may be objected that these restrictions should go into the object-language of the calculus rather than into its semantical meta-language. We should then have the Barcan-type formula '$\forall x(\Box^n A)\rightarrow\Box^n\forall x(A)$' as an axiom-schema, rather than the stronger thesis '$\Box^n A\rightarrow\Box^n\forall x(A)$'. But it seems to me that the restrictions may reasonably be regarded as belonging quite properly to the semantical meta-language. They simple constitute a special case of the general principle that each kind of symbol-occurrence in the calculus should be reserved, in the intended interpretation, for the performance of just one logical function, and just one kind of symbol-occurrence for each function. Unbound individual letters are to function just as references to individuals, predicate letters are to function just as predicates and not also as references to individuals, and so on. It is also in keeping with this general principle, for example, that in order to avoid familiar paradoxes about quantification into non-extensional contexts, we may conveniently require non-extensionality-generating expressions like 'necessarily', or 'it is logically true that', to be represented only by the modal operators of the calculus and not to form any part of expressions represented by predicate letters.[2] Nor may any predicate letter represent assertions of existence, i.e. membership of the domain of discourse: only the existential quantifier may do this. Otherwise a counter-example may arise to our analogue for the converse of the

[1] *The Logic of Scientific Discovery*, pp. 68 f.

[2] Cf. *The Diversity of Meaning*, p. 213

Barcan formula, viz. $\Box^n \forall x(A) \rightarrow \forall x(\Box^n A)$.[1] Finally, since we already have a non-denumerable infinity of primitive modal operators as logical constants, we do not need out modal operator variables to be capable, when unbound, of designating individual modalities. Every formula containing one or more unbound modal operator variables can therefore be made provably equivalent to its closure for these variables: i.e. in addition to $\forall \Box^x(\Box^x A) \rightarrow \Box^x A$ we shall have $\Box^x A \rightarrow \forall \Box^x(\Box^x A)$.

It may well be that some useful concepts of evidential support do not conform to the principle represented by $\Box^n A \rightarrow \Box^n \forall x(A)$ where x is free in A. But we should hardly want to exclude that law from applying when it is assumed, as is certainly assumed by the Instantial Comparability Condition (3), that all questions about the extent of support for singular predictions should be treated as cases of questions about support for the general predictions which these predictions instantiate. If we assume the uniformity of nature in regard to evidential comparisons, then our assumption is authorised by the principle that gives us $\Box^n A \rightarrow \Box^n \forall x(A)$. Indeed, it is a consequence of this principle that, if the truth of Qx is physically necessary, then $\forall x(Qx)$ is a law of nature: every causation of one property of an event by another instantiates some pattern of natural uniformity.[2]

IX

I now set out the calculus $O4^4$ in detail, together with a few comments on its theorems.

Primitive Symbols:

Monadic, dyadic, etc., predicate letters: represented in the syntactical meta-language by ' P ' or ' Q ' or ' R ' or ' S '.

Individual variables: represented by ' x ' or ' y ' or ' z '.

---

[1] Cf. S. A. Kripke, 'Semantical Considerations on Modal Logic', *Act. Phil. Fenn.* xvi, p. 90

[2] I hope to show elsewhere how $O4^4$ can also be interpreted as generalised deontic logic, with a representation of the so-called principle of ethical generalisation or universalisability. This interpretation will produce a representation of the concept of 'better than' that is applicable to cases of conflicting obligations as well as just to cases where obligations conflict with morally indifferent or prohibited courses of action or morally indifferent courses conflict with prohibited ones: contrast L. Aqvist, 'Deontic Logic Based on a Logic of "Better"', *Act. Phil. Fenn.* xvi, pp. 285 ff.

Modal operators, viz. squares differentiated by superscripts, one of which denotes some designated real number greater than zero and the others of which denote distinct real numbers less than this designated number but greater than zero. ' $\Box^{d}$ ' represents the square with the superscript that denotes the designated real number, and ' *d* ' represents this superscript. ' $\Box^{c}$ ' represents some particular square with a superscript denoting a real number that is less than or the same as the designated one, and ' *c* ' represents this superscript. ' $\Box^{m}$ ' and ' $\Box^{n}$ ' represent any primitive squares, unless otherwise stated, and ' *m* ' and ' *n* ', respectively, represent the superscripts of the squares so represented. Expressions connecting the letters ' *c* ' ,' *d* ', ' *m* ', or ' *n* ' with one another by signs of equality or inequality will be used to state, not mention, restrictions on the superscripts these letters represent in terms of equalities or inequalities between the numbers denoted by these superscripts.

Modal variables, viz. squares with letters as superscripts, represented by ' $\Box^{x}$ ' or ' $\Box^{y}$ ' or ' $\Box^{z}$ '.

Brackets, dash, ampersand and quantifier: represented by ' ( ) ', ' – ', ' & ', and ' $\exists$ ', respectively.

Individual and modal variables are also represented indifferently by ' *v* ' and ' *w* '; and numerical subscripts will be used to generate additional metalinguistic symbols where necessary.

A *well-formed formula* (wff) is defined recursively:

$Px_1x_2 \ldots x_j$ is a wff, where P is j-adic. If A and B are wff, so are $-A$, $(A \& B)$, $\exists v(A)$, $\Box^{n}A$ and $\Box^{x}A$.

Hereafter ' A ', ' B ', ' C ', ' D ', ' E ', ' F ', ' G ', ' H ', ' I ', and ' J ' represent wff.

An occurrence of *v* in A is a *bound* occurrence if in a wf part of A of the form $\exists v(B)$. Otherwise it is free.

Definition-schemata:

(Definition-schemata are statements in the syntactical metalanguage that describe permissible abbreviations of wff and wf parts of wff. In each case the new meta-linguistic symbolism used should be taken to represent symbolism of similar shape to itself.)

1. $(A \rightarrow B)$ for $-(A \& -B)$
2. $(A \leftrightarrow B)$ for $((A \rightarrow B) \& (B \rightarrow A))$
3. $(A \vee B)$ for $(-A \rightarrow B)$
4. $(A \rightarrow^{n} B)$ for $\Box^{n}(A \rightarrow B)$

5. $(A \leftrightarrow^{n} B)$ for $((A \rightarrow^{n} B) \& (B \rightarrow^{n} A))$
6. $\Diamond^{n} A$ for $-\Box^{n} - A$
7. $\Box^{\circ} A$ for $(\exists \Box^{x} (\Box^{x} A) \vee - \exists \Box^{x} (\Box^{x} A))$
8. $(A \rightarrow^{\circ} B)$ for $\Box^{\circ} - (A \& - B)$
9. $\forall v(A)$ for $-\exists v(-A)$
10. $c[H, E] > 0$ for $\exists \Box^{x} (E \rightarrow^{c} \Box^{x} H)$
11. $c[H, E] = 0$ for $-c[H, E] > 0$
12. $c[H, E] \geqslant 0$ for $(c[H, E] > 0 \vee c[H, E] = 0)$
13. $c[H, E] \geqslant n$ for $(E \rightarrow^{c} \Box^{n} H)$
14. $c[H, E] < n$ for $-c[H, E] \geqslant n$
15. $c[H, E] \leqslant n$ for $\forall \Box^{x} ((E \rightarrow^{c} \Box^{x} H) \rightarrow (\Box^{n} H \rightarrow^{d} \Box^{x} H))$
16. $c[H, E] > n$ for $-c[H, E] \leqslant n$
17. $c[H, E] = n$ for $(c[H, E] \geqslant n \ \& \ c[H, E] \leqslant n)$
18. $c[H, E] \geqslant c[I, F]$ for $\forall \Box^{x} ((F \rightarrow^{c} \Box^{x} I) \rightarrow (E \rightarrow^{c} \Box^{x} H))$
19. $c[H, E] > c[I, F]$ for $-c[I, F] \geqslant c[H, E]$
20. $c[H, E] = c[I, F]$ for $(c[H, E] \geqslant c[I, F] \ \& \ c[I, F] \geqslant c[H, E])$

Hereafter outermost brackets will always be omitted, and other brackets also where convenient.

A *meta-theorem* is a statement of the form ' A is a theorem ', which may be abbreviated '$\vdash A$'. The criteria of theoremhood are as follows:

21. If A is a truth-functional tautology, then $\vdash A$.
22. If $\vdash A$ and $\vdash A \rightarrow B$, then $\vdash B$.
23. If $\vdash A$, then $\vdash \Box^{d} A$ and $\vdash \Diamond^{n} A$ where $c > n$.
24. $\vdash (A \rightarrow^{n} B) \rightarrow (\Box^{n} A \rightarrow^{n} \Box^{n} B)$ where $n \geqslant c$.
25. $\vdash (A \rightarrow^{n} B) \rightarrow (\Box^{n} A \rightarrow \Box^{n} B)$ where $c > n$.
26. $\vdash \Box^{n} A \rightarrow \Box^{m} A$ where $n > m$.
27. $\vdash \Box^{c} A \rightarrow A$.
28. $\vdash \Box^{n} A \rightarrow \Box^{n} \forall x(A)$.
29. $\vdash A \rightarrow \forall v(A)$ where v is not free in A or v is $\Box^{x}$.
30. $\vdash \forall v(A) \rightarrow B$ where B is like A except for having free occurrences of x, or occurrences of $\Box^{n}$, wherever A has free occurrences of v.
31. $\vdash \forall v(A \rightarrow B) \rightarrow (\forall v(A) \rightarrow \forall v(B))$.
32. If $\vdash A$, then $\vdash \forall x(B)$, where B is like A except for having free occurrences of x wherever A has free occurrences of y.
33. If $\vdash A$ and $\Box^{n} E$ occurs in A, and if, for some places in which $\Box^{n} E$ occurs in A, and for all m, $\vdash B$ where B is like A except for having $\Box^{m} E$ in these places, then $\vdash C$ where C is like A except for having free occurrences of $\Box^{x}$ in these places. (That is, if *M* is a meta-theorem, and O is a meta-linguistic symbol representing a modal,

operator, and O occurs in one or more places in M without restriction as to the superscript of the operator it represents, then N is a meta-theorem, if N is like M except for having the representation of a free occurrence of a modal variable wherever M has O.)

The *interpretation* envisaged for wff of $O4^4$ rests on the usual formal-logical replacements for brackets, dash and ampersand. Restrictions on the replacement of predicate letters and free individual variables, and replacements for the modal operators, have been already proposed (in VII and VIII above). Suitable replacements for $\exists x$ and for bound occurrences of individual variables have been proposed elsewhere, along with restrictions designed to avoid semantical antinomies.[1] The initial $\exists\Box^x$ in each wf part $\exists\Box^x(A)$ of a wff is to be replaced by ' there is a quantity of reasons such that ', and free occurrences of $\Box^x$ in the wf part A of $\exists\Box^x(A)$ by ' there exists at least this quantity of the reasons that would justify anyone's accepting it to be true that . . .', where the blank '. . .' is to be filled by whatever replaces A. If $\Box^x$ occurs freely in a wff A, A is to have the same replacement as that for the closure of A in respect of modal variables.

Meta-Theorems: (Most of the proofs are quite straight-forward and are omitted for the sake of brevity.)

We first derive a few useful auxiliary rules:

34. If $\vdash A\&B$, then $\vdash A$ and $\vdash B$.
35. If $\vdash A$ and $\vdash B$, then $\vdash A\&B$.
36. If $\vdash A_1\rightarrow A_2$, $\vdash A_2\rightarrow A_3$, . . . $\vdash A_{n-1}\rightarrow A_n$, then $\vdash A_1\rightarrow A_n$.
37. If $\vdash A\rightarrow B$, then $\vdash\Box^n A\rightarrow\Box^n B$.
37A. If $\vdash A$, then $\vdash\Box^n A$.

Certain important properties belong to a restricted class of modal operators, e.g.:

38. $\vdash\Box^n A\rightarrow A$ where $n>c$
39. $\vdash(A\,\&(A\rightarrow^n B))\rightarrow B$ where $n>c$
40. $\vdash A\rightarrow\Diamond^n A$ where $n>c$

This law is one of the reasons why we cannot use $\Diamond^n A$ where $n>c$, instead of $\Box^n A$ where $c>n$, to represent statements of partial support for A. Another reason is that no associative law, like 50 below, is provable for $\Diamond^n A$, and 50 is needed to prove the General Conjunction Condition, 135.

---

[1] *The Diversity of Meaning*, pp. 200 and 219 f.

Next certain standard properties of the modal operators emerge:

41. $\vdash(A\rightarrow^{n}B)\rightarrow(\Box^{m}A\rightarrow\Box^{m}B)$ where $n>m$
42. $\vdash(\Box^{n}A\&(A\rightarrow^{n}B))\rightarrow\Box^{n}B$
43. $\vdash(A\rightarrow^{n}B)\rightarrow(-B\rightarrow^{n}-A)$
44. $\vdash((A\rightarrow^{n}B)\&(B\rightarrow^{n}C))\rightarrow(A\rightarrow^{n}C)$
45. $\vdash((A\rightarrow^{n}B)\&(A\rightarrow^{n}C))\rightarrow(A\rightarrow^{n}(B\&C))$
46. $\vdash\Box^{n}A\leftrightarrow(-A\rightarrow^{n}A)$
47. $\vdash\Box^{n}A\rightarrow(B\rightarrow^{n}A)$
48. $\vdash\Box^{n}-A\rightarrow(A\rightarrow^{n}B)$
49. $\vdash\Box^{n}(A\&B)\rightarrow(\Box^{n}A\&\Box^{n}B)$
50. $\vdash(\Box^{n}A\&\Box^{m}B)\rightarrow\Box^{m}(A\&B)$ where $n>m$
  i. $(\Box^{n}A\&\Box^{m}B)\rightarrow(\Box^{m}A\&\Box^{m}B)$ where $n>m$ 26, 21, 22
  ii. $(\Box^{m}A\&\Box^{m}B)\rightarrow(-(A\&B)\rightarrow^{m}A)\&(-(A\&B)\rightarrow^{m}B))$ 47, 47, 35, 45, 22
  iii. $(\Box^{n}A\&\Box^{m}B)\rightarrow(-(A\&B)\rightarrow^{m}(A\&B))$ where $n>m$ i, ii, 45, 36
  iv. $(\Box^{n}A\&\Box^{m}B)\rightarrow\Box^{m}(A\&B)$ where $n>m$ iii, 46, 34, 36
51. $\vdash(\Box^{n}A\&\Box^{n}B)\leftrightarrow\Box^{n}(A\&B)$
52. $\vdash\Box^{n}A\rightarrow\Box^{n}(AvB)$
53. $\vdash(\Box^{n}Av\Box^{m}B)\rightarrow\Box^{m}(AvB)$ where $n>m$
54. $\vdash(A\&-\Box^{n}A)\rightarrow-\Box^{n}(A\&-\Box^{n}A)$
55. $\vdash\Diamond^{m}A\rightarrow\Diamond^{n}A$ where $n>m$
56. $\vdash-\Box^{n}(A\&-A)$
  i. $-\Box^{n}--(A\&-A)$ where $c>n$ 21, 23
  ii. $-\Box^{n}--(A\&-A)$ where $n>c$ 21, 40, 22
  iii. $-\Box^{n}--B\rightarrow-\Box^{n}B$ 21, 37, 21, 22
  iv. $-\Box^{n}(A\&-A)$ i, iii, 22; ii, iii, 22

That is, there can be no reasons (no objectively valid reasons, no evidential support on balance in the universe as a whole) in favour of a self-contradictory statement.

57. $\vdash-(\Box^{n}A\&\Box^{m}-A)$ 56, 50, 35, 21, 22
58. $\vdash-\Box^{n}Av-\Box^{m}-A$
59. $\vdash\Box^{n}A\rightarrow\Diamond^{m}A$

That is if the total evidence is favourable to A, it is not favourable to $-A$: cf. the Negation Condition for confirmation functors, proved in 86 below with the aid of 57.

60. $\vdash\Box^{n}(A\&-A)\rightarrow B$
61. $\vdash(\Box^{n}A\&\Box^{n}-A)\rightarrow B$

62. $\vdash(\Box^n A \to^d \Box^m A) \vee (\Box^m B \to^d \Box^n B)$
i. $-(\Box^n A \to^d \Box^m A) \to (\Box^m B \to^d \Box^n B)$ where $m > n$ 26, 23, 21, 22
ii. $-(\Box^n A \to^d \Box^m A) \to (\Box^m B \to^d \Box^n B)$ where $n > m$ 26, 23, 21, 22
iii. $(\Box^n A \to^d \Box^m A) \vee (\Box^m B \to^d \Box^n B)$ i, ii, 3

This law is needed for the proof of 98 below, and thus for the proof (*ad fin.*) that the degrees of evidential support denoted by the confirmation-functors of $O4^4$ form a chain, or simply ordered set.

Laws of iteration apply to a restricted class of modal operators:

63. $\vdash\Box^n A \to \Box^n \Box^m A$ where $n > m > c$
i. $\Box^n A \to ((B \vee - B) \to^n A)$ 21, 37
ii. $\Box^n A \to^n \Box^m A$ where $n > m$ 26, 37A
iii. $((B \vee - B) \to^n A) \to ((\Box^n (B \vee - B) \to^n \Box^n A) \& (\Box^n A \to^n \Box^m A)$ where $n > m > c$ ii, 24, 35, 21, 22
iv. $((B \vee - B) \to^n A) \to (\Box^n (B \vee - B) \to^n \Box^m A)$ where $n > m > c$ iii, 44, 36
v. $(\Box^n (B \vee - B) \to^n \Box^m A) \to (\Box^n \Box^n (B \vee - B) \to \Box^n \Box^m A)$ where $n > m > c$ 24, 26, 27, 36
vi. $(\Box^n A \to ((\Box^n \Box^n (B \vee - B) \to \Box^n \Box^m A)$ where $n > m > c$ i, iv, v, 36
vii. $\Box^n \Box^n (B \vee - B)$ 21, 37A, 37A
viii. $\Box^n A \to \Box^n \Box^m A$ where $n > m > c$ vi, vii, 35, 21, 22

64. $\vdash\Box^n \Box^m A \to \Box^n A$ where $m > c$
65. $\vdash\Diamond^n \Box^m A \to \Diamond^n A$ where $m > c$

$\Box^\circ$ has certain properties that it shares with all or some of the primitive modal operators, e.g.:

66. $\vdash(A \to^\circ B) \to (\Box^\circ A \to \Box^\circ B)$
67. $\vdash(\Box^\circ A \& (A \to^\circ B)) \to \Box^\circ B$
68. $\vdash((A \to^\circ B) \& (B \to^\circ C)) \to (A \to^\circ C)$
69. $\vdash\Box^\circ A \to (B \to^\circ A)$
70. $\vdash\Box^\circ (A \ \& \ B) \leftrightarrow (\Box^\circ A \ \& \ \Box^\circ B)$
71. $\vdash\Box^\circ A \leftrightarrow \Box^\circ \Box^\circ A$

$\Box^\circ$ also has certain unique properties, e.g.:

72. $\vdash\Box^\circ A$
73. $\vdash\Box^\circ A \ \& \ \Box^\circ - A$
74. $\vdash\Box^\circ (A \ \& \ - A)$
75. $\vdash\Box^n A \to \Box^\circ A$

The standard properties of quantifiers emerge, e.g.:

76. $\vdash\forall v(A \ \& \ B) \leftrightarrow (\forall v(A) \ \& \ \forall v(B))$

77. $\vdash \forall v(A \to B) \to (\exists v(A) \to B)$ where v is not free in B

77A. $\vdash \forall v(-A) \to -\exists v(A)$

77B. $\vdash \forall v(A \to B) \to (\exists v(A) \to \exists v(B))$

There are also special laws for quantification in modal contexts, e.g.:

78. $\vdash \Box^{n} A \to \forall x(A)$ where $n > c$ 28, 26, 27, 36

79. $\vdash \forall x(\Box^{n} A) \to \Box^{n} \forall x(A)$
   i. $\forall x(\Box^{n} A) \to \forall x(\Box^{n} \forall x(A))$ 28, 32, 31, 22
   ii. $\forall x(\Box^{n} A) \to \Box^{n} \forall x(A)$ i, 30, 36

80. $\vdash \Box^{n} \forall x(A) \to \forall x(\Box^{n} A)$
   i. $\forall x(\Box^{n} \forall x(A) \to \Box^{n} A)$ 30, 37, 32
   ii. $\Box^{n} \forall x(A) \to \forall x(\Box^{n} A)$ i, 31, 22, 29, 36

81. $\vdash \exists x(\Box^{n} A) \to \Box^{n} A$
   i. $\exists x(\Box^{n} A) \to \Box^{n} \forall x(A)$ 28, 32, 77, 22
   ii. $\exists x(\Box^{n} A) \to \Box^{n} A$ 30, 37, i, 36

82. $\vdash \Diamond^{n} A \to \forall x(\Diamond^{n} A)$ (also proved from 28)

83. $\vdash \Diamond^{n} \forall x(A) \to \forall x(\Diamond^{n} A)$
   i. $\Box^{n} - A \to \Box^{n} - - \exists x(-A)$ 30, 21, 22, 37
   ii. $\exists x(- - \Box^{n} - A) \to \Box^{n} - - \exists x(-A)$ 21, i, 36, 32, 77, 22
   iii. $\Diamond^{n} \forall x(A) \to \forall x(\Diamond^{n} A)$ ii, 21, 22, 6, 9

84. Let an occurrence of C be an *m*th grade occurrence in A, where $d > m$, if and only if it (i) forms a proper part of the occurrence in A of a wf part of A of the form $\Box^{m}$B, and (ii) does not form a proper part of the occurrence in A of a wf part of A of the form $\Box^{n}$E, where $n > m$, or of the form $\Box^{x}$E. An occurrence of C is a *d*th grade occurrence in A if and only if it forms a proper part of the occurrence in A of a wf part of A of the form $\Box^{d}$B or $\Box^{x}$B. Then if C, D, B are such that B results from A by the substitution of D for zero or more *n*th grade occurrences of C in A, and $x_1, x_2, \dots x_k$ is a complete list of individual variables occurring freely in C and D, then $\vdash \forall x_1 \forall x_2 \dots \forall x_k\ (C \leftrightarrow^{m} D \to)(A \leftrightarrow B)$ where $m > n$. That is, evidential support operators generate characteristic hierarchies of non-extensional substitutivity.

We are now in a position to explore the logical syntax of confirmation-functors, developing first some theorems about zero degree of evidence:

85. $\vdash c[H, E] > 0$ 21, 10, 11, 12

86. $\vdash \Diamond^{c} E \to (c[H, E] > 0 \to c[-H, E] = 0)$
   i. $(E \to^{c} (\Box^{n} H \ \& \ \Box^{n} - H)) \to \Box^{c} - E$ 57, 21, 22, 37
   ii. $(E \to^{c} \Box^{n} H) \to (\Diamond^{c} E \to -(E \to^{c} \Box^{n} - H))$ 45, i, 36, 21, 22, 6

iii. $(E \rightarrow^{c} \Box^{x} H) \rightarrow \forall \Box^{x}(\Diamond^{c} E \rightarrow -(E \rightarrow^{c} \Box^{x} - H))$

ii, 33, 29, 22 ; 31, 22; 29, 36

iv. $\forall \Box^{x}(\Diamond^{c} E) \rightarrow ((E \rightarrow^{c} \Box^{x} H) \rightarrow \forall \Box^{x}(-(E \rightarrow^{c} \Box^{x} - H)))$

iii, 31, 36, 21, 22

v. $\forall \Box^{x}(\Diamond^{c} E) \rightarrow (\exists \Box^{x}(E \rightarrow^{c} \Box^{x} H) \rightarrow \exists \Box^{x}(\forall \Box^{x}(-(E \rightarrow^{c} \Box^{x} - H))))$

iv, 29, 22; 31, 22; 77B, 29, 36

vi. $\exists \Box^{x}(\forall \Box^{x}(-(E \rightarrow^{c} \Box^{x} - H))) \rightarrow - \exists \Box^{x}(E \rightarrow^{c} \Box^{x} - H)$

29, 21, 22; 77A, 36

vii. $\forall \Box^{x}(\Diamond^{c} E) \rightarrow (\exists \Box^{x}(E \rightarrow^{c} \Box^{x} H) \rightarrow - \exists \Box^{x}(E \rightarrow^{c} \Box^{x} - H))$

v, vi, 35, 21, 22

viii. $\Diamond^{c} E \rightarrow (c[H, E] > 0 \rightarrow c[-H, E] = 0$ 29, vii, 36, 10, 11

This Negation Condition for confirmation-functors in eliminative induction may be contrasted with Carnap's[1] T 59—1p, viz. $c(-h, e) = 1 - c(h, e)$.

87. $\vdash \Diamond^{c} E \rightarrow (c[H, E] = 0 \vee c[-H, E] = 0)$
88. $\vdash \Diamond^{c} E \rightarrow (c[H, E] = c[-H, E] \rightarrow c[H, E] = 0)$. Cf. the very different law that follows from Carnap's T 59—1p, viz. if $c(h, e) = c(-h, e)$ then $c(h, e) = \frac{1}{2}$.
89. $\vdash \Diamond^{c} E \rightarrow (c[H, E] > c[-H, E] \rightarrow c[-H, E] = 0)$
90. $\vdash \Diamond^{c} E \rightarrow (c[H, E] \geqslant c[-H, E] \rightarrow c[-H, E] = 0)$
91. $\vdash c[H, E] = 0 \rightarrow c[H \& I, E] = 0$. Compare Carnap's T 59—6a, viz. if $c(h, e) = 0$ then $c(h.\,i, e) = 0$.
92. $\vdash c[H \vee I, E) = 0 \rightarrow c[H, E] = 0$. Compare Carnap's T 59—6s, viz. if $c(h \vee i, e) = 0$ then $c(h, e) = 0$.

Next we develop some general principles in the logic of statements about non-zero degrees of evidential support:—

93. $\vdash c(H, E) \geqslant 0 \;\&\; c[H, E] \leqslant d$. Compare Carnap's T 59—1a, viz. $0 \leqslant c(h, e) \leqslant 1$.
94. $\vdash [c\, H, \Box^{n} H] \geqslant n$
95. $\vdash c[H, E] \geqslant n \rightarrow c[H, E] > 0$
96. $\vdash c[H, E] > c[I, E] \rightarrow c[H, E] > 0$
97. $\vdash c[H, E] \geqslant n \vee c[H, E] \leqslant n$. With the aid of 62 we first prove $((E \rightarrow^{c} \Box^{m} H) \rightarrow (\Box^{m} H \rightarrow^{d} \Box^{n} H)) \vee ((E \rightarrow^{c} \Box^{m} H) \rightarrow (\Box^{n} H \rightarrow^{d} \Box^{m} H))$.
98. $\vdash c[H, E] \geqslant c[I, F] \vee c[I, F] \geqslant c[H, E]$. Also proved with the aid of 62.
99. $\vdash (E \;\&\; c[H, E] \geqslant n) \rightarrow \Box^{n} H$. This kind of detachment is made possible by defining confirmation-functors in the object-language: contrast Carnap's meta-linguistic functors.

---

[1] *Logical Foundations of Probability* (1950), pp. 315 ff.

Some general laws of transitivity are provable for comparisons between different degrees of evidential support. For example, for the relation of 'greater than or equal to' we have:

100. $\vdash (c[H, E] \boldsymbol{>} c[I, F] \ \& \ c[I, F] \boldsymbol{>} c[J, G]) \rightarrow c[H, E] \boldsymbol{>} c[J, G]$

101. $\vdash [(c[H, E] \boldsymbol{>} n \ \& \ c[I, F] \boldsymbol{<} n) \rightarrow c[H, E] \boldsymbol{>} c[I, F]$

102. $\vdash (c[H, E] \boldsymbol{>} c[I, F] \ \& \ c[I, F] \boldsymbol{>} n) \rightarrow c[H, E] \boldsymbol{>} n$

103. $\vdash (c[H, E] \boldsymbol{<} n \ \& \ c[H, E] \boldsymbol{>} c[I, F]) \rightarrow c[I, F] \boldsymbol{<} n$

For the relation of 'greater than' we have corresponding laws:

104. $\vdash (c[H, E] > c[I, F] \ \& \ c[I, F] > c[J, G]) \rightarrow c[H, E] > c[J, G]$

105. $\vdash (c[H, E] > n \ \& \ c[I, F] < n) \rightarrow c[H, E] > c[I, F]$. The proof uses 62.

106. $\vdash (c[H, E] > c[I, F] \ \& \ c[I, F] > n) \rightarrow c[H, E] > n$

107. $\vdash (c[H, E] <^{n} \ \& \ c[H, E] > c[I, F]) \rightarrow c[I, F] < n$

So too for the relation of 'equal to':

108. $\vdash (c[H, E] = c[I, F] \ \& \ c[I, F] = c[J, G]) \rightarrow c[H, E] = c[J, G]$

109. $\vdash (c[H, E] = n \ \& \ c[I, F] = n) \rightarrow c[H, E] = c[I, F]$

110. $\vdash (c[H, E] = c[I, F] \ \& \ c[I, F] = n) \rightarrow c[H, E] = n$

It is also possible to prove 102, 106, 109 and 110 with '0' in place of 'n'.

We are only concerned with the syntax of confirmation-functors so far as they relate, directly or indirectly, to the strength of support by eliminative induction for universal generalisations. Positive support is thus always ultimately derived from favourable responses to tests of certain kinds on universal hypotheses *plus the absence of unfavourable responses* (Evidential Relevance Condition). Hence, if E would support H in this way, and F would support —H, E and F cannot both be true:

111. $\vdash (c[H, E] > 0 \ \& \ c[-H, F] > 0) \rightarrow \Box^{c} -(E \ \& \ F)$. Compare 131 as another example of a law that depends for its *raison d'être* on the Evidential Relevance Condition.

Under certain special conditions for H or E there are restrictions on the degree of evidential support that E can give, or not give, to H:

112. $\vdash \Box^{n} H \rightarrow c[H, E] \boldsymbol{>} n$ where $n \boldsymbol{>} c$. Provable from 63 and 47. Compare Carnap's T 59—1c, viz. if h is L-true, $c(h, e) = 1$.

113. $\vdash \Box^{n} H \rightarrow c[H, E] > 0$ where $n \boldsymbol{>} c$. Provable from 112 and 95.

114. $\vdash \Box^{n} -H \rightarrow (\Diamond^{c} E \rightarrow c[H, E] = 0)$ where $n \boldsymbol{>} c$. Provable from 113 and 86. We may compare Carnap's T 59—1f, viz. if h is L-false, $c(h, e) = 0$.

115. $\vdash\Box^{n}E\rightarrow(c[H, E]>0\rightarrow c[H, F]>0)$ where $n\geqslant c$. Provable with the help of 47. That is, if a statement is at least as sound as our confirmation-metric, it is too sound to provide any evidential support in the normal way: it can give no support to a hypothesis unless that hypothesis is supported by any statement whatever. Compare where $c = d$, Carnap's notion (p. 289) of null confirmation on tautological evidence.

116. $\vdash\Box^{n}-E\rightarrow c[H, E]\geqslant n$ where $n\geqslant c$. That is, if the denial of a statement is at least as sound as our confirmation-metric, the statement may with impunity be said to support anything whatever. Compare Carnap's argument (pp. 295 f.) that if any value at all is to be assigned to c[H, E] where E is L-false we should normally then have $c[H, E] = 1$.

Some relationships between H and E also impose restrictions:

117. $\vdash(E\rightarrow^{n}H)\rightarrow c[H, \Box^{n}E]\geqslant n$ where $n\geqslant c$

117A. $\vdash c[H, E]\geqslant c\rightarrow(E\rightarrow^{c}H)$

118. $\vdash\Diamond^{c}\Box^{c}E\rightarrow((E\rightarrow^{n}-H)\rightarrow c[H, E] = 0)$ where $n\geqslant c$

i. $(\Box^{c}E\rightarrow(\Box^{m}H \;\&\; \Box^{c}-H))\rightarrow-\Box^{c}E$ 57, 21, 22

ii. $\Diamond^{c}\Box^{c}E\rightarrow-(\Box^{c}E\rightarrow^{c}(\Box^{m}H\&\Box^{c}-H))$ i, 37, 21, 22, 6

iii. $\Diamond^{c}\Box^{c}E\rightarrow((\Box^{c}E\rightarrow^{c}\Box^{c}-H)\rightarrow-(\Box^{c}E\rightarrow^{c}\Box^{m}H))$ ii, 45, 35, 21, 22

iv. $\Diamond^{c}\Box^{c}E\rightarrow((E\rightarrow^{n}-H)\rightarrow-(\Box^{c}E\rightarrow^{c}\Box^{m}H))$ where $n\geqslant c$ 26, 24, 36; iii, 35; 21, 22

v. $(E\rightarrow^{c}\Box^{m}H)\rightarrow(\Box^{c}E\rightarrow^{c}\Box^{m}H)$ 27, 21, 22, 37a, 37

vi. $\Diamond^{c}\Box^{c}E\rightarrow((E\rightarrow^{n}-H)\rightarrow-(E\rightarrow^{c}\Box^{m}H)$ where $n\geqslant c$ iv, v, 35, 21, 22

vii. $\Diamond^{c}\Box^{c}E\rightarrow((E\rightarrow^{n}-H)\rightarrow c[H, E] = 0$ where $n\geqslant c$ vi, 33, 29, 22; 31, 22; 29, 35; 77A, 35; 21, 22

The restriction described in 117 is relatively unimportant compared with Carnap's T 59—1b, viz., if $\vdash e\rightarrow h$, $c(h, e) = 1$, since the domain of our confirmation-theory is not the interval between E's contradicting a universal hypothesis U and E's implying U, but rather the interval between E's implying there to be no support at all for U and E's implying U to state a necessity. With 118, however, we may compare Carnap's T 59—1e, viz. if $\vdash e\rightarrow -h$, $c(h, e) = 0$. The main difference from Carnap's law (viz. the condition $\Diamond^{c}\Box^{c}E$ in 118) is due to the fact that in eliminative induction our counter-evidence to U—the variety of situation that falsifies U—must itself be a potential law of nature or a case or consequence of one: theoretically, no hypothesis is eliminated except in favour of another.

The relationship between two hypotheses may impose much more important restrictions, e.g.:

119. $\vdash(H\rightarrow^{d}I)\rightarrow(c[I, E]\geqslant c[H, E])$. This is what I have called the Consequence Condition: cf. Carnap's T 59—2d, viz. if $\vdash h_1\rightarrow h_2$, or $\vdash e.h_1\rightarrow h_2$, then $c(h_1, e)\leqslant c(h_2, e)$. The proof of 119 leans on the characteristic modal thesis 37, and with the aid of 119 we may derive an equivalence condition for hypotheses:
120. $\vdash(H\leftrightarrow^{d}I)\rightarrow c[H, E] = c[I, E]$. Compare Carnap's T 59—1i, viz. if $\vdash h_1\leftrightarrow h_2$, $c(h_1, e) = c(h_2, e)$.

Certain corresponding conditions also hold, viz.:

121. $\vdash c[H\rightarrow I, E]=d\rightarrow c[I, E]\geqslant c[H, E]$. Compare Carnap's T 59—6f, viz. if $c(h_1\rightarrow h_2, e) = 1$, $c(h_1, e)\leqslant c(h_2, e)$.
122. $\vdash c[H\leftrightarrow I, E]=d\rightarrow c[H, E]=c[I, E]$. Compare Carnap's T 59—6g, viz. if $c(h_1\leftrightarrow h_2, e) = 1$, $c(h_1, e) = c(h_2, e)$.

As detachments from the Consequence Condition (119) we have:

123. $\vdash c[H\rightarrow I, E]\geqslant c[I, E]$
124. $\vdash c[H\rightarrow I, E]\geqslant c[H, E]$
125. $\vdash c[H, E]\geqslant c[H\&I, E]$. Compare Carnap's T 59—2e, viz. $c(h.i, e)\leqslant c(h, e)$.
126. $\vdash c[H \vee I, E]\geqslant c[H, E]$. Compare Carnap's T 59—2f, viz. $c(h, e)\leqslant c(h\vee i, e)$.

Corresponding to 125 we also have:

127. $\vdash c[H\&I, E) = n\rightarrow c[H, E]\geqslant n$
128. $\vdash c[H\&I, E]>n\rightarrow c[H, E]>n$

Similarly, corresponding to 126, we have:

129. $\vdash c[H, E] = n\rightarrow c[H \vee I, E]\geqslant n$
130. $\vdash c[H, E]>n\rightarrow c[H \vee I, E]>n$

It is also possible to prove 127, 128, 129 and 130 with '0' for '*n*'.

A consequence condition for evidential statements is derivable:

131. $\vdash(E\rightarrow^{n}F)\rightarrow c[H, E]\geqslant c[H, F]$ where $n\geqslant c$. If H is a universal hypothesis, then, on pain of irrelevance to H, both E and F must (at least implicitly) either report a counter-instance to H or to some other universal statement I from which H is logically deducible or deny that one has emerged from such-and-such tests, the results of which include all the available evidence for H or I (Evidential Relevance Condition). Hence, where E implies F, it cannot be the case that F is relevant to H while E

is not. Instead either E and F are both irrelevant to H, or E is and F is not, or both conform to the requirements of the Evidential Relevance Condition, and each of these cases validates 131. If H is not a universal hypothesis, neither E nor F can give it positive support except through some universal hypothesis thus related to them: cf. e.g. 134 below.

From 131 we obtain an equivalence condition for evidential statements:

132. $\vdash(E\leftrightarrow^{n}F)\rightarrow c[H, E] = c[H, F]$ where $n \geqslant c$. Compare Carnap's T 59—1h, viz. if $\vdash e_1\leftrightarrow e_2$, $c(h, e) = c(h, e_2)$.

With the help of 28 and 119 we can derive certain theorems about the relations between support for universal statements and support for their instantiations, e.g.:

133. $\vdash c[\forall x_1 \forall x_2 \ldots \forall x_k\ H, E] = c[A, E]$ where A differs from H just in having y_1 where H has x_1, y_2 where H has x_2, … y_k where H has x_k.

134. $\vdash c[\forall x_1 \forall x_2 \ldots \forall x_k\ H, E] > c[\forall x_1 \forall x_2 \ldots \forall x_k\ I, F] \leftrightarrow c[A, E] > c[B, F]$ where as in 133 and B differs from I analogously in having z_1 where I has x_1, etc.

This version of the Instantial Comparability Condition is, like 133, unrestricted in respect of its evidential statements. E.g. it is possible for E to refer to the element denoted by y_1, and F to that denoted by z_1, contrary to what is required by (3). But, once it is understood that we are formalising here a concept of evidential support that treats all such support as being essentially inductive and mediated through generalisations it becomes clear that in fact the Instantial Comparability Condition for eliminative induction requires no restriction of this nature in respect of its evidential statements. As emerges in 117, even if B and F in 134 are identical, what we can infer from this identity is the value of $c[B, \Box^{c}F]$—not of $c[B, F]$.

With the help of the associative law for modal operators, 50, we can easily prove the General Conjunction Condition for confirmation-functors:

135. $\vdash c[H, E] \geqslant c[I, E] \rightarrow c[H\&I, E] = c[I, E]$. Contrast Carnap's T 59—1n, viz. $c(h.i, e) = c(h, e) \times c(i, e.h)$.

135A. $\vdash c[H\&I, E] = c[I, E] \rightarrow c[H, E] \geqslant c[I, E]$.

Corresponding to 135 we have:

136. $\vdash(c[H, E] \geqslant n\ \&\ c[I, E] = n) \rightarrow c[H\&I, E] = n$

137. $\vdash(c[H, E] \geqslant 0\ \&\ c[I, E] = 0) \rightarrow c[H\&I, E] = 0$.

With the help of 135 and 119 we obtain laws of detachment for hypotheses, e.g.:

138. $\vdash c[(H\rightarrow I), E] > c[H, E] \rightarrow c[I, E] > c[H, E]$
138A. $\vdash c[H, E] > c[H\rightarrow I, E] \rightarrow c[H, E] > c[I, E]$
139. $\vdash c[(H\rightarrow I), E] > c[H, E] \rightarrow (c[H, E] > n \rightarrow c[I, E] > n)$.

With the help of 133 and 135 we can prove the Instantial Conjunction Condition:

140. $\vdash c[A_1 \& A_2 \& \ldots \& A_k, E] = c[A_1, E]$ where A_i differs from A_j in just that A_i has x_1 wherever A_j has y_1, x_2 wherever A_j has $y_2, \ldots x_r$ wherever A_j has y_r.

With the help of 133 we can also prove that if evidence is irrelevant to a universal hypothesis it is also irrelevant to any substitution-instances of that hypothesis, and vice versa, viz.:

141. $c[\forall x_1 . \forall x_2 \ldots \forall x_k H, E\&F] = c[\forall x_1 \forall x_2 \ldots \forall x_k H, E] \leftrightarrow c[A, E\&F] = c[A, E]$ where as in 133.

Finally, it can easily be shown that the confirmation-functors contextually defined in $O4^4$ satisfy standard axioms for expressions denoting the members of a simply ordered set, or chain, such as the positive integers, ordered with respect to magnitude, or the real numbers. (This is perhaps hardly surprising in view of the fact that the superscripts of the primitive modal operators denote real numbers.)

Birkoff[1] gives the following axioms for a chain:

P1 : for all x, $x > x$.
P2 : if $x > y$ and $y > x$, then $x = y$.
P3 : if $x > y$ and $y > z$, then $x > z$.
P4 : given x and y, either $x > y$ or $y > x$.

Writing c[H, E] for x, c[I, F] for y, and c[J, G] for z, we have P3 in meta-theorem 100, and P4 in 98. For P1 and P2, respectively, we have

142. $c[H, E] > c[H, E]$ 21, 29, 22, 18
143. $(c[H, E] > c[I, F] \;\&\; c[I, F] > c[H, E]) \rightarrow c[H, E] = c[I, F]$ 21, 20

The Queen's College,
Oxford

---

[1] G. Birkhoff, *Lattice Theory* (1948)

Brit. J. Phil. Sci. (1966), *Volume* 17, *Number* 2, 127–140
Printed in Great Britain

THE USE OF STATISTICS IN EXPLANATION*

ARTHUR W. COLLINS

THE idea of a pattern of argument called 'statistical explanation' has been developed as the complement of the analysis of the logic of explanation variously known as the deductive model, the covering-law model, and the nomological model. As deductive explanations consist in subsuming what is explained under a general law, so statistical explanations are supposed to 'subsume' what they explain under a statistical generalisation. From this viewpoint, the 'logic of statistical explanation' proposes to clarify the structure and the conditions for the acceptability of such statistical explanatory arguments. I think this viewpoint is incorrect, and in this paper I show what is wrong with the preliminary thoughts on which the idea of a logic of statistical explanation is based, and I state briefly the kinds of relationships in which the information formulated in statistical sentences may be expected to be of help when something needs explaining.

'Explanation' will be used to cover only explanations of particular occurrences, in contrast with natural uses of 'explanation' that include coverage of explanations of laws, regularities and statistical correlations, and also explanations of such other sorts of things as mathematical proofs and obscure lines of poetry.

I give only a brief exposition of the projected logic of statistical explanation and none at all of the deductive analysis from which it derives. For both, I rely on, without trying to summarise, the many detailed accounts in the works of Professor C. G. Hempel: or the logic of statistical explanation, 'Deductive Nomological *v.* Statistical Explanation,'[1] and for the deductive model, the classic paper on which P. Oppenheim collaborated with Hempel, 'Studies in the Logic of Explanation'.[2]

It is important to emphasise that the roots of the concept of statistical explanation lie in the deductive or 'covering-law' model.

---

* Received 21.ix.65

[1] *Minnesota Studies in the Philosophy of Science*, Minneapolis, 1963, **3**, 98–169

[2] *Philosophy of Science*, 1948, **15**, 135–175, also reprinted in part in *Readings in the Philosophy of Science*, Feigl and Brodbeck (ed.), New York, 1953

Hempel and Oppenheim included in their presentation of the deductive model this remark:

> The concept of causal explanation which has been examined here, is capable of various generalizations. One of these consists in permitting [the premisses] . . . to include statistical laws. . . . [The other is the acceptance of inductive rather than only deductive inference in explanatory argument.] Both of these extensions open important prospects and raise a variety of new problems. In the present essay, however, these issues will not be further pursued.[1]

It is these issues that are pursued in 'Deductive Nomological *v.* Statistical Explanation'. Here, following a new exposition of the deductive theory, Hempel introduces the statistical model (systematisation) for explanatory argument with the help of this preliminary comparison:

> This [deductive] argument is basically of the form:
>
> (A)[2] All F are G.
> x is F.
> x is G.
>
> The form of the statistical explanation, on the other hand, appears to be expressible as follows:
>
> (B)[2] Almost all F are G.
> x is F.
> x is almost certain to be G.[3]

A statistical explanation, then, is an explanation wherein a statistical generalisation serves as an approximation to a general law. While we cannot say that all F are G, at least we can say that most are. From this point of view, a statistical explanation has a certain irreducible inferiority. We would always prefer a universal generalisation. And the worth of a statistical explanation is measured, in one dimension at least, by the degree to which the statistical generalisation approaches the universal. 'Most' is not as good as 'almost all', and 'the majority' is pretty much useless. This is apparent in light of another view that ordinarily accompanies deductivism, namely, the assertion of the structural identity of arguments that make possible explanations and predictions. Plainly, 'Some F are G and x is F' would be slim support for the prediction', 'x is G', while, 'Virtually all F are G', would be virtually as good as 'All F are G'.

---

[1] *Readings in the Philosophy of Science*, p. 350 [2] My indices [3] p. 125

I think all of this is wrong but I will come to that below. The bulk of Hempel's technical effort in the statistical part of 'Deductive Nomological *v.* Statistical Explanation' consists in manoeuvres to outflank 'the ambiguity of statistical explanation' which could lead to explanatory disasters: 'Briefly, it [the ambiguity] consists in the fact that if the explanatory use of a statistical generalisation is conceived in the manner of [(B), above], then a statistical explanation of a particular event can, in general, be matched by another one, equally of the form [(B)], which statistically explains the non-occurrence of the same event.'[1] Diverging from this article, but remaining faithful to its spirit, the gist of the difficulty and its solution is the following: although (B) is satisfied, it may turn out that x is an H as well as being and F, and it may be true that almost no H are G even though almost all F are G, and this, even if almost all or all H are F. To schematize

(C) Almost all F are G.
x is F.
But, x is also H.
And, very few H are G.

Therefore, it is not the case that x is almost certain to be G, even though x is F.
On the contrary, x is almost certain to be non-G.

And, therefore, x's being F cannot be understood to explain x's being G.

The substance of this problem is familiar in inductive logic and that is where Hempel turns for a solution. Suppose we can say how much inductive support, how much probability, a certain hypothesis derives from certain premisses, statistical or otherwise. We still cannot say how probable the hypothesis is *simpliciter*, or how we should guide our behaviour with respect to it, because other true premisses may vitiate the support of those we have considered. Before we can speak of the probability of a hypothesis by itself we must be sure that we have taken into consideration all of the relevant evidence and this, in the absence of a viable criterion of relevance, includes all our knowledge. Thus, as one criterion for the evidential adequacy of the premisses for a statistical explanation, Hempel adduces 'the requirement of total evidence'.[2]

---

[1] My indices, p. 127

[2] This phrase is Carnap's from *The Logical Foundations of Probability*, Chicago, 1950

The requirement of total evidence will become important to our diagnosis of the logic of statistical explanation below. For the moment we should note one thought without which the requirement would make no sense. That is the thought that

(D) If very few F are G then x's being F cannot explain x's being G.

(D) is tacitly accepted in the logic of statistical explanation. Only strong statistical relations can support explanation. This view seems to be thought too obvious to argue, but its obviousness depends on analogues with the deductive-model theory of explanation, not on explanatory practice. What is obvious is only that, ' Very few F are G ', is a poor approximation to a universal law and of no use whatever for the prediction that x is G. In order to demonstrate that the logic of statistical explanation is on the wrong track we need only note that (D) is quite false rather than obviously true; that the danger for which technical remedies are sought (the requirement of total evidence in ' Deductive Nomological *v.* Statistical Explanation ') is not a danger and, more generally, that the relevance of ' P per cent of F are G ', to an explanation of x's being G is independent of the size of P. Consider the following illustration: A small number of people who are injected with penicillin develop an allergic skin reaction. I invent a figure: 2 per cent. Suppose x develops an allergic skin reaction. Why did he? Well, by hypothesis, because he received an injection of penicillin. Does this story make sense? Is the explanatory claim illogical? No, the explanation would be accepted and rightly accepted in most situations in which an explanations for x's getting a rash is sought.

There are a few points about this that ought to be insisted on. First, in ordinary cases in which we know why someone got an allergic reaction, i.e. from the injection, we do not have any other information with which we could strengthen the correlation between injection and rash. Second, the explanation we give may be wrong and we may or may not come to find that out. But our confidence in the explanation would ordinarily and justifiably be very high. So far, I have not said why it would be or should be high but think, ' What percentage of these explanations will turn out to be wrong? ' For any explanation, the possibility that it may be wrong cannot be entirely removed. Note that in the real situation our confidence in this explanation could not be understood as a function of the 2 per cent relationship, as though we would be slightly more confident if it were 3 per cent. Third, there are contexts in which this explanation would not be what was wanted and would not be illuminating because it was already known and a need for explanation survived. This, in particular, would be true of contexts in which the reasons for allergic reactions to penicillin is at issue. But there are always such contexts for any explanation, and the existence of such contexts does not show that we are wrong or just guessing when we agree that x got a rash as a consequence of the penicillin injection.

Now, following the format of statistical explanation, we have:

(E) Very few of the penicillin-injected are rash-getters.
x is penicillin-injected.
———————————
x is almost certain not to be a rash-getter.

But x is a rash-getter and is a rash-getter because x is penicillin-injected. (E) is a straightforward counterinstance to (D). (E) cannot be the ground for accepting the explanation of x's rash but neither can it be the ground for excluding the explanation. Therefore, we should expect that the prospect of finding such a weak statistical relationship cannot be the threat to explanatory hypotheses that statistical explanation theory takes it to be: the threat for which the requirement of total evidence is introduced as a defence. In this expectation we are right.

Suppose 98 per cent of cases of the bite of a certain snake are fatal. And suppose that x, the fatality perplexing us, was bitten by that sort of snake and then died. This is satisfaction of the preliminary statistical schematisation (B) and would support a prediction since, on the basis of these premisses, x is almost certain to be a fatality. But caution is in order since x may also be an H such that very few H's succumb even to this snake's bite. This would render the first statistic irrelevant and scotch the snake hypothesis. But consider: x immediately had the antidote and it is also true that hardly any of the bitten die if they get the antidote promptly. Therefore, on the basis of this new evidence, which the requirement of total evidence would be sure not to miss, x would have been considered almost certain not to be a fatality. This means that the danger that was foreseen has become a reality, but it turns out not to exclude the explanation at all. x died of snake bite. He was one of those rare cases that the antidote did not save. We may wonder why the antidote did not work on x: Was it a bad batch? but that is not to wonder whether we are right to cite the bite. And again, maybe the explanation is wrong and x died of a heart attack.

CALCUTTA CENTRAL

The point is that the fact that he died in spite of getting the antidote which usually works is a negligible reason for thinking the explanation is wrong. We may wonder why the antidote does not work on a few people including x. This is like wondering why penicillin only affects some people that way. Such wondering presupposes that it is the penicillin that affects those few, that reference to the penicillin *explains*.

This shows, once again, that the logic of statistical explanation fails to get at the substance of explanations to which the information formulable in statistical propositions is relevant. ' Very few F are G ', is no approximation to a law, but, contrary to the logic of statistical explanation, being an F can explain being a G. And even if the worst that the requirement of total evidence is designed to prevent happens, the explanations presumed to be rejected need suffer nothing.

If I were a deductivist and a statistical-explanation theorist I think I would regard all this as a reason to plunge in again to make a new model and more rules for the logic of statistical explanation. Reconsideration of preliminary notions will lead us in another direction. I want to introduce a distinction between two kinds of statistical proposition. The distinction has nothing to do with the strength of correlations but has much to do with the relevance of statistics to explanation. I do not deceive myself about the approximate character of this distinction. It holds up well enough for the uses for which I have framed it. The two kinds of statistical statements will have almost self-explanatory names: first, *pure statistical propositions*, and second, *potential statistical causal assessments*. Both can be expressed by formally identical sentences. Thus,

Pure statistical propositions:

25 per cent of the draws from a normal deck of cards are spades.
51 per cent of human births are male.
Over 3 per cent of sequences of five tosses of a normal coin are all heads.

Potential statistical causal assessments:

25 per cent of the picknickers who ate the mayonnaise became ill.
51 per cent of those opposed to the proposal were property owners.
Over 3 per cent of those injected develop a rash.

Taking the form of all statistical propositions to be, ' P per cent of F are G ', a correlation is a potential statistical causal assessment if, whether or not it would be true, it would make sense to say that being F is the cause, a cause, a causal factor, part of the cause, etc, of being G; or that G is the cause, a cause, etc. of being F. For example, it is intelligible to wonder whether, or to assert that, having had the mayonnaise was the cause of having become ill and, if we do assert this, '25 per cent of those who ate mayonnaise became ill', is a causal assessment. Therefore, it is a potential causal assessment even if we do not. In contrast, it makes no sense to wonder whether or to assert that being a draw from a normal deck is the cause of being a spade draw. There are other significant aspects of the distinction between pure statistics and causal assessments. Only for the latter is it natural to think of improving a correlation, or think of one correlation as better than another. For instance, 'We can improve the correlation if we consider only those who had the mayonnaise after 3 p.m.', or, 'But there is an even better correlation between the liverwurst and subsequent illness'. On the other hand, what could one mean by saying that 'One-sixth of the rolls of a die are aces', is a rather poor correlation and looking for ways to improve it? This could be a joke made by someone who proposed to load the die. A second significant factor accompanying the distinction is that the converses of statistical causal assessments tend to be interesting or useful, while the converses of pure statistics are useless or absurd. '99 per cent of those who became ill had eaten the mayonnaise', would further implicate the mayonnaise. '99 per cent of all spade draws are draws from normal decks', is a useless, if an intelligible statistic. The converses of casual assessments would always be relevant but not always actually of use. Thus, '1 per cent of those who develop rashes like that rash were injected with penicillin', would not help the explanation of a particular rash that cites the penicillin, but would not hinder it either.

What I want to say with the help of this distinction is simple: Only potential statistical causal assessments have any relevance to explanations of individual occurrences. Only if we actually think of a factor as a causal factor would we advance it in an explanation. When we do so explain, the explanatory force of what we say derives from the fact that we assert that something caused what we explain and not from the strength of the statistical correlation, although correlations and their specific magnitudes *may be* what lead us to think of a factor as a cause.

All this is almost true. I am forced to be guarded by the diversity of perplexities that can beset us. Suppose someone tosses a coin five times and gets five heads. Suppose he is astonished. We might mitigate his astonishment by telling him that over 3 per cent of such sequences are all heads. Not having reflected he had thought of such an outcome as much less likely. Whether by extension, analogy, or in its own right, I think all this might be presented as ' an explanation of what happened ', and here it would be a pure statistical correlation that explains. Note this residual difference: the explanation here is not condensable as, ' You got five heads because it was a sequence of five tosses of a normal coin ', paralleling, ' x got the rash because he was injected with penicillin '. In this case there is something ironically suggestive of the deductivist's identification of explanatory and predictive argument. In a margnial way the statistic shows that the outcome was predictable, that we should not have been so sure it would not happen, not because it was five tosses of a normal coin but because it does happen 3 per cent of the time. Note also that a low correlation may explain something in this sort of context where a high correlation would not. It is hard to think of any context for this: Why did x draw a card other than the spade ace? Well, almost all draws are other than the spade ace.' Anyone who understood enough to ask the question would have to know that much, so the question suggests a causal query (Why did the card trick not work?) and the answer is empty because it adduces no causal factor. Cases involving pure statistical correlations are marginally explanatory, require thinking of special contexts and, to understate, are far from paradigmatic of the explanation of particular occurrences.

Let us consider normal explanations to which statistics are relevant, namely, explanations in which the relevant statistical proposition is a causal assessment. We need to try to understand how these explanations manage to explain in spite of very low correlations, and that comes to trying to understand how we can think of a factor F as a causal factor for G even though very few F are G. This is not very difficult to understand and, in fact, becomes puzzling only when we take the viewpoint of deductive-model theorists who conceive divergence from the universal generalisation to be the measure of imperfection of an explanation. This orientation of thought has led statistical-explanation theory to be as inattentive to real statistical investigation as it has been shown to be inattentive to real explanatory practice. Someone who wanted to know whether the mayonnaise was responsible for the illnesses would not be satisfied with, ' 25 per cent of those who ate the mayonnaise became ill ', nor would he be satisfied if it were 100 per cent. At least, he would want that information required for Mill's method of concommitant variation. That is, he would want to know the percentage of F that are G, of G that are F, of non-F that are G, and of non-G that are F. He would want to know what percentage of all members of the domain of discourse (here, all the picnickers) are F, and what, G. This he would interpret in the light of correlations for other possible causal factors: the liverwurst. And, were he careful, he would worry about the possibility that the real cause might be some factor that happens accidentally to be co-extensive with being F: insecticide on the mayonnaise spoon. Is not this something like the requirement of total evidence? Perhaps it is. The difference is that such an investigation is not premissed upon, and does not seek either to preserve or to discover any statistic to the effect that a large percentage of F are G. This kind of statistical investigation produces explanations for x being G without worrying about the magnitude of P in, ' P per cent of F are G ', and often the correctness of such explanations is a moral certainty. Think, for instance, of medical diagnoses. Many diagnoses are uncertain, many are entirely certain. We do not need to argue this in any event since we can more easily explain the power of causal explanations in simpler circumstances where the causal factor has a very low correlation with the event it explains and where, if explicit statistical formulations exist, they are not required to support the explanation.

What makes us think that x is G because it is F when very few F are G? I apologise for giving a detailed answer to this question since the answer is a common sense answer. Nonetheless, the following considerations have become the sort of which philosophers need to be reminded, so far have they been blown off course by the winds of theory. In cases where we make this kind of causal estimate we reason more or less like this: ' It does not matter that few F are G. The important thing to focus on is that, after all, x is G.' We think having had penicillin explains a man's rash although penicillin seldom produces that rash. This is sensible because it if was not the penicillin it must have been something else, and the likelihood that other factors were present that might have produced the rash is not great. If we rule out other things that could have made x be G, our confidence grows without any corresponding growth in the strength of the correlation between F and G. Knowing that the event we want to explain *has occurred* and that this sort of event customarily only comes about in one of certain roughly known ways, we are pretty sure about an explanation as soon as we discover that conditions for one of those ways for getting x to be G obtained just before x got to be G. This is brought out in the contrast: While only 2 per cent of the penicillin-injected are rash-getters, more than 99 per cent of those injected by penicillin who do get a rash get it as a result of the injection. The latter statistical relationship shows us how often we will be right explaining as we do. It does not reconstruct any explanatory argument. Perhaps even more striking are those cases in which we know that being G can only come about in one way. Then we come to know that x was F only by inference from the fact that it is G. Such explanatory contexts are not at all rare. The reader will supply the explanations for particular occurrences that would be based on the following:

> Only a beaver makes a dam like that.
> This kind of rheumatic inflammation is always part of the aftermath of a streptococcus infection.
> Only a glacier leaves this stratification.
> Only Aunt Millie knocks like that.

For such formulas, it is irrelevant in what percentage of cases and under what circumstances a beaver does make such a dam, or Aunt Millie does knock. If there were any such figures, the correlations might be tiny. In many situations where we need an explanation, whether we get one or not depends on the relative rareness of the sort of occurrence in which we are interested. If it is rare and we have some understanding of it, an explanation will probably be easy. How easy it is to give a very likely explanation of one case of plague when there is only one other case in the country. In contrast, to explain one case of 'flu we must be satisfied with, ' There is a lot of it around '.

Why should anyone overlook these considerations and propose near-universality as a significant ground for an explanatory relationship? Our reasoning, above, made use of information that is inaccessible to statistical-explanation and deductive-explanation theorists, information that they cannot accept and still retain the general shape of their theories. This information is the fact that the event we want to explain has occurred. Recall the deductivist view that the sort of information that would support a really illuminating explanation would also have made possible a prediction of the occurrence explained had that information been available previously. The technical counterpart of this idea lies in the forms presumed to make explicit the logic of explanatory argument. The occurrence of the event to be explained or the likelihood of its occurrence are the conclusions of the arguments, (A) and (B), with which an explanation is alleged to be explicitly reconstructed. The decisive point, then, is that if we include the fact that x is G as a premiss, the conclusion ' x is G ' would follow trivially and the whole reconstruction of the argument would collapse. Looking at rather than theorising about explanations has shown that it should collapse.

As philosophers we can sympathise with others who hold to theoretical convictions more or less in the face of the facts. As I have said, a factor is explanatory only if it is causal. But we are all empiricists. We find no causal connection *a priori*, and *a posteriori*, we find no causal nexus. This story is familiar. We never see F getting G to come about. They just do happen, one after the other. Why do we say F causes G? Because they always happen one after the other (deductive model), or almost always (statistical model). As we have seen this is not a very good account of the way we actually assess causal relationships. Far from the fields where explaining is done, however, our empiricist convictions concerning causality seem to sustain a powerful principle, and this principle plays a crucial behind-the-scenes role in the logic of explanation. That is the principle that if the preceding events, the conditions and circumstances are exactly the same, then the same thing always happens. Therefore, had we the means to describe the situation well enough we could say that in such and such situation G always happens, we could always predict G, and everything causal and explanatory that could rightly be said about G would be said in our description.

I have not introduced this thought either to endorse it or to rebut it. It gets at something we all seem to think, but it has a characteristic philosophical wildness too. I think it is definitely a conceptual prop for the deductive theory of explanation but its strength in that connection will not be tested here. Far from supporting the main doctrines of the logic of statistical explanation this basic principle concerning causality is fundamentally incompatible with the conception of statistical explanation we have examined.

Let us review the considerations that led statistical-explanation theory to the requirement of total evidence. This requirement was to defend against the danger that there might be a factor H which discriminates a subclass of F such that individuals that are F and H are rarely G. The principle of causality cited would assure us that there is such an H and, therefore, the conjunction of the requirement of total evidence and this principle of causality would rule out every statistical explanation. For consider, if almost all F are G then, according to the principle of causality, there are better descriptions F′ and F″ such that all F′ are G and all F″ are non-G. Being F″, if only we could discover this feature, would explain an individual's being non-G. But to be F″ is just to be F with some discriminating factor, and H, that is the threat of which led to the requirement of total evidence. Therefore, whenever almost all F are G, being F would be an ingredient in explanations of individual instances of non-G as well as instances of G. In our relative ignorance, we would not be able to describe using the discriminating factor H and, in consequence, would not be able to produce any explanation for being non-G that cite being F and H. Nonetheless, insofar as we believe that the same conditions always produce the same result (our principle of causality), we believe that there is such an H that would make possible such explanations. We could accept a statistical explanation then, only by winking at evidence which we know to exist but have not yet got.

With the help of a technical device Professor Hempel achieves a much simpler statement of some of these relationships. His objectives, of course, are not ours. Applied to our context, he would introduce a class defined as all those individuals that either (i) are the individual x, or (ii) are describable as F″. Let this class be designated K. Then, following Hempel's example, we could say that whenever almost all F are G and x is F, it is also the case that almost no K are G and x is K. And our point is that the existence of such a class K is, on the one hand, guaranteed by the principle of causality and, on the other hand, suffices for the exclusion of an explanation for being G that cites being F. I have disregarded this neater means of expression, heretofore, although it makes for a persuasive parallelism, lest I give the impression that the difficulty here is only technical, for, the definition of the class K does not correspond to any intuitive way of dividing thing up. I want to emphasize that the problems here are substantive, and arise because the concepts of explanation and cause have been deformed in the effort to produce a logic of statistical explanation.

The same difficulty for candidate statistical explanations can be formulated using the familiar schemata for inductive logic. Let *h* be ‘ x is G ’, and *e* be ‘ Almost all F are G and x is F ’. Then ‘ P(*h*, *e*) is very high ’, reads, ‘ *h* has a very high inductive probability on the basis of the evidence *e* ’. The reconstruction of the explanatory argument for x's being G now has this form:

(i) P(*h*, *e*) is very high, where *e* is the total known evidence, therefore, *h* is very likely.

I use the word ' likely ' here for the categorical ascription of probability, allowed once the requirement of total evidence is satisfied, in contrast with the relational ascription of ' probable ' only meaningful with reference to specific evidence. The results we have obtained mean that whenever we have an argument satisfying (i) we believe that there is undiscovered evidence *f*: ' Almost all (F and K) are non-G and x is K ', rendering the following argument equally valid:

(ii) P(*h*, *e* and *f*) is very low, where *e* and *f* is the theoretical total evidence, therefore *h* is very unlikely.

Of course, if we did come to possess this theoretically existent evidence, we would be able to explain x's being G using the deductive model.[1] But, given a causal interpretation of the statistical relationships involved, every statistical explanation must be judged unacceptable within the framework of statistical explanation theory.

What this means is that the commitment of theorists of statistical explanation to the conceptions of the deductive theory of explanation is so great that the purported statistical extension fails to admit any explanations at all. For the only explanations that would satisfy the requirement of total evidence, if we include evidence guaranteed by the principle of causality would be deductive explanations. A statistical explanation was supposed, from the point of view we have considered, to be valued as an approximation to a law-covered, deductive explanation. Hence, the emphasis on high percentages. However, if only almost all F are G, some are not. Abstraction from the contexts in which we actually explain appears to give force to this line of thought: F is not (quite) sufficient for G. Given an x that is F, will it be G? Some factor other than F must be decisive, but then this other factor must also be the decisive ingredient in an explanation of x being G. If this abstract viewpoint were right, appeal to any less-than-universal generalisation would just show that we did not yet know why x is G. This is the source of the ' ambiguity of statistical schematisation ' at the intuitive level.

---

[1] Supposing, that is, that the deductive model is a satisfactory account of the explanatory power of some explanations of particular occurrences. The correctness of this supposition lies outside the present context. I do not think the deductive model is at all acceptable and, while it is not a premiss for any arguments here, this thought is closely bound up with my view of statistical explanation. See my, ' Explanation and Causality ', *Mind*, forthcoming.

If we do not construe a statistical explanation as a causal assessment, it fails to be explanatory (excepting the marginal explanatory value of pure statistics). If we do construe a statistical correlation as indicating a cause, the requirement of total evidence will always reject the explanation of an occurrence because a perfectly parallel argument for non-occurrence from other true premisses will be constructible in principle if not available in fact. This should not be thought a paradox, nor even particularly surprising. It is the natural outcome of characterising the acceptability of explanations by reference to formal features: the form of premisses, magnitude of correlations, degree of inductive support, and deducibility of the event to be explained. The first and ineliminable criterion for the adequacy of an explanation must be based on what it does for a man who wants an explanation. This depends on contextual factors that are not reflected in the forms of propositions or the structure of inferences. The logic of explanation has ordinarily been approached as if ' the explanation of an event ' were in important ways like ' the factor of a number ' or ' the translation of a word '. For these, it is the man who is wrong if he does not like what we give him. It may be helpful, instead, to consider the explanation of an event to be in important ways like ' the description of an object '. Then the temptation to erect formal criteria may be lessened.

Department of Philosophy
Columbia University

Brit. J. Phil. Sci. (1966), *Volume* 17, *Number* 2, 141–143
Printed in Great Britain

NOTE

A Paradox of Zero Information

THE present note is to be read as a continuation of my earlier note 'A Comment on Miller's New Paradox of Information',[1] which I will assume to be known. Thus I shall not explain here again the terminology and symbolism used there—unless, of course, I deviate from it.

In that earlier note I suggested in section I (and, in a slightly specialized form, again in section IV) that the Straight Rule may be reformed by stating it as one of the following two very plausible *conditional rules*, (Rc) and (Rcr):

(Rc) If E^a_r is true, then $P(A) = r$.

(Here '$P(A)$' might *look* at first sight like an '*absolute probability*', or what Carnap calls a 'measure function', although it is not so intended; see below.)

(Rcr) If E^a_r is true, then $P(A, E^a_r) = r$.

(Here '$P(A, E^a_r)$' is clearly a relative probability (the probability of A, *given* E^a_r).)

Now I wish first to make clear that the apparently 'absolute' form (Rc) will be satisfactory only if we do *not* interpret '$P(A)$' as a genuine *absolute* probability, or 'measure function', but rather as the probability (or acceptability, or degree of confirmation, or what not) *at the time* at which the accumulated evidence is stated by E^a_r. In other words, we shall have to replace '$P(A)$' in (Rc) by something like '$P_d(A)$', to be read 'The Probability (or 'degree of confirmation' etc.) of A at the date d'.

Obviously, we shall have to introduce a corresponding variable d into our evidence statement. In order to avoid further complications to our notation by additional subscripts, I propose the following re-interpretation of our symbolism for evidence statements:

(Convention) The symbol 'D^a_r' (or similarly 'E^a_r') is to be used as a metalinguistic name of the evidence statement

$$fr(a,d) = r \text{ (or } fr(a,e) = \text{r)}$$

or in words, 'The observed frequency at the date d (or e) of the property a equals r'.

This suggests the following *prima facie* definition of '$P_d(A) = q$':

(*Prima Facie* Definition of P_d)

$$P_d(A) = q \;\equiv\; (x) \text{ (if } D^a_x \text{ is true, then } P(A, D^a_x) = q)$$

---

[1] This *Journal*, 1966, 17 (65), 61

I shall not argue for the *prima facie* adequacy of this definition, especially as it will later turn out to be inadequate. (Incidentally, if the definiens is replaced by the expression

$$(\mathrm{E}x)(D^a_x \text{ is true and } P(A, D^a_x) = q)$$

then the definition turns out to be just as inadequate, and for precisely the same reasons.)

Now if we substitute in (Rc) the definiens of this *Prima Facie* Definition for the definiendum, we find the simple result: rule (Rc) entails rule (Rcr). But this means that if (Rcr) leads to a contradiction, then (Rc) must lead to a contradiction too. Thus we can now concentrate our criticism upon (Rcr), in the form:

(Rcr) $\qquad$ If D^a_r is true then $P(A, D^a_r) = r$.

That this rule is subject to a special form of Miller's paradox may be seen as follows:

We remember that 'D^a_r' (according to our 'Convention' stated above) is the name of

$$fr(a,d) = r$$

or in words 'the relative frequency of a, as observed up to the date d, equals r'. It follows that if we substitute '$fr(a,d)$' for 'r', we obtain a tautology: our information shrinks to *zero*. We thus have

(1) $\qquad \vdash D^a_{fr(a,d)}$

(2) $\qquad \vdash E^a_{fr(a,e)}$

(3) $\qquad \vdash D^a_{fr(a,d)} \equiv E^a_{fr(a,e)}$

Now in view of (1) and (2) we obtain from (Rcr) by the rule of detachment

(4) $\qquad P(A, D^a_{fr(a,d)}) = fr(a,d) \qquad$ by (Rcr) and 1

(5) $\qquad P(A, E^a_{fr(a,e)}) = fr(a,e) \qquad$ by (Rcr) and 2

and by the principle that logically equivalent evidence statements may be substituted for each other,

(6) $\qquad fr(a,d) = fr(a,e) \qquad$ by 3, 4 and 5

But (6) is, obviously, paradoxical; for it is to be interpreted as saying that for *any* two dates, d and e, the frequencies of a observed up to these two dates are equal (and therefore, equal to the *a priori* absolute probability of a).

Thus our 'reform'—the 'way out' of adopting a conditional straight rule—is ineffective.

Incidentally, it can be shown, by the same argument, that our *Prima Facie* Definition of $P_d(A)$ leads alone—without (Rc)—to an analogous contradiction; for we obtain by instantiation, or substitution for 'x' '$fr(a,d)$',

(7) $\qquad P_d(A) = q \supset P(A, D^a_{fr(a,d)}) = q$

Now in view of the general principle of the probability calculus

(8) If $\vdash B$, then $P(A) = P(A,B)$

where '$P(A)$' denotes the absolute probability (or measure) of A, we obtain from (7),

(9) $P_d(A) = q \supset P(A) = q$ by 7, 1, and 8

and therefore, by substitution of 'e' for 'd':

(10) $P_e(A) = q \supset P(A) = q$ by 9.

But (9) and (10) lead to a contradiction. For by substituting '$P_d(A)$' for 'q' in (9) and '$P_e(A)$' for q in (10) we get

(11) $P_d(A) = P_e(A)$, by 9 and 10

the same paradoxical result as before. Of course, this merely shows that our *Prima Facie* Definition of P_d is unacceptable as it stands.

I may mention that there is still a way out of all this: we can introduce a further *ad hoc* reform, of (Rcr) as well as of the *Prima Facie* Definition, and thus of (Rc). It consists in replacing (wherever 'D^a_r is true' occurs as an implicans) 'If D^a_r is true' by 'If D^a_r is *Te*' where '*Te*' means true *and* empirical, i.e. synthetic. (However, this may not be quite enough: '*Te*' may have to be interpreted in an even stronger sense, as meaning 'true and synthetic *as well as* being our total evidence, in Carnap's sense'.)

However, a look at the way this new case of Miller's paradox arises shows that, formally considered, it depends only on the introduction of a new variable (or parameter), 'd'. Now it is clear that 'd' may be interpreted in '$fr(a,d)$' in various ways: not only as a date, but also, for example, as the (total) empirical knowledge of a certain person (say, Daniel) at a certain date. Thus we might read '$fr(a,d)$' as 'the observed frequency, according to Daniel's (total) empirical knowledge'. But this would make the introduction of (total) empirical knowledge in the new *ad hoc* reform counter-intuitive.

Another trouble with this new *ad hoc* reform is that in the case of zero information P_d ought automatically to coincide with absolute probability. But the possibility of this is blocked by the new reform.

I do not doubt, nevertheless, that some inductivists will, undeterred, seize upon this latest reform and declare that it is not *ad hoc* but in perfect agreement with their inductive intuition—unless indeed they prefer to ignore the paradox altogether as they have done before with inconvenient arguments (some of which are mentioned in Miller's and my earlier notes).

University of London KARL R. POPPER

DISCUSSIONS

Miller's So-called Paradox of Information

Miller claims to deduce contradictory conclusions from the assumption which he writes as

$$P\left(A, E\frac{a}{r}\right) = r \qquad (1)$$

which on one interpretation becomes a form of the 'straight rule' (Reichenbach's rule of induction). That is, it may be read as follows :

The logical probability that an individual event of kind a will occur, relative to the evidence that the so-far observed frequency of events of kind a is r, is r.

For example, the logical probability that this die will fall 5-uppermost at its next throw, relative to the evidence that the frequency of 5-uppermost fallings among throws of this die is $\frac{1}{6}$, is $\frac{1}{6}$.

Miller concludes that this rule must be rejected. However, his argument contains a logical fallacy. To expose this fallacy, let us state his proof a little more fully, making explicit the necessary instantiations and generalisations.

(I shall write 'P (a at next throw/the frequency of a's $= r$)' for 'The logical probability that a will occur at the next throw relative to the evidence that the so-far observed frequency of events of kind a is r'.)

1. For all r, P (a at next throw/the frequency of a's $= r$) $= r$—this is the assumption to be refuted by deducing a contradiction from it.
2. P (a at next throw/the frequency of a's $= \frac{1}{2}$) $= \frac{1}{2}$
   —from 1 by U.I., substituting '$\frac{1}{2}$' for 'r'.
3. P (a at next throw/the frequency of a's $=$ the frequency of not-a's) $=$ the frequency of not-a's
   —from 1 by U.I., substituting 'the frequency of not-a's' for 'r'.
4. P (a at next throw/the frequency of a's $= \frac{1}{2}$) $= P$ (a at next throw/ the frequency of a's $=$ the frequency of not-a's)
   —because the frequency of a's $= \frac{1}{2}$ iff the frequency of a's $=$ the frequency of not-a's, and—as Sir Karl Popper explains in his Comment—in the probability calculus if $B = C$, then $P(A, B) = P(A, C)$.
5. The frequency of not-a's $= \frac{1}{2}$
   —from 2, 3, and 4, by the transitivity of '$=$'.
6. For all frequencies of a's, the frequency of not-a's $= \frac{1}{2}$
   —from 5 by U.G.
7. Where the frequency of a's $= \frac{1}{6}$, the frequency of not-a's $= \frac{1}{2}$
   —from 6 by U.I.
8. Where the frequency of a's $= \frac{1}{6}$, the frequency of not-a's $= \frac{5}{6}$
   —by arithmetic.

Since the frequency of a's may $= \frac{1}{6}$, 7 and 8 are contradictory, and 1 must be rejected.

This may be compared with the following parallel argument, in which a similar contradiction is deduced from an undeniable algebraic statement, namely, 'For all x, the value of a, given that $a = x$, $= x$'. (I shall write '$V(a/a = x)$' for 'The value of a, given that $a = x$'.)

1. For all x, $V(a/a = x) = x$
2. $V(a/a = \frac{1}{2}) = \frac{1}{2}$
   —from 1 by U.I.
3. $V(a/a = 1\text{-}a) = 1\text{-}a$
   —from 1 by U.I.
4. $V(a/a = \frac{1}{2}) = V(a/a = 1\text{-}a)$
   —because $a = \frac{1}{2}$ iff $a = 1\text{-}a$, and if $B \equiv C$, then the value of a, given B, is equal to the value of a, given C.
5. $1\text{-}a = \frac{1}{2}$
   —from 2, 3, and 4, by the transitivity of '$=$'
6. For all values of a, $1\text{-}a = \frac{1}{2}$
   —from 5 by U.G.
7. Where $a = \frac{1}{6}$, $1\text{-}a = \frac{1}{2}$
   —from 6 by U.I.
8. Where $\mathrm{a} = \frac{1}{6}$, $1\text{-}a = \frac{5}{6}$
   —by arithmetic

Since 7 and 8 are contradictory, 1 must be rejected!

The fallacy, in each argument, occurs at step 6. In the parallel argument, step 5 has been derived from step 3, and the 'a' in step 3 has the *particular* value assigned to it by '$a = 1\text{-}a$'. In step 3, 'a' is *not* a free variable, it is *not* implicitly universally quantified 'for all values of a'. Consequently 'a' in step 5 is not implicitly so quantified either, it is not the sort of 'arbitrarily selected individual' to which U.G. can be applied; consequently to apply U.G. at step 6, to introduce the explicit quantification 'for all values of a', is fallacious. Similarly in the original argument, in step 3 'the frequency of not-a's' is not implicitly universally quantified 'for all frequencies of a's', but has the *particular* value assigned to it by 'the frequency of a's = the frequency of not-a's'. Consequently the same applies to step 5, since it is derived from step 3, and the universal generalisation at step 6, explicitly introducing this quantification, is equally fallacious.

Since in each case the final contradiction has been deduced invalidly, the acceptance of step 1 would not commit us to this contradiction. Miller's 'paradox', therefore, does not tell against the form of the 'straight rule' in question, any more than the parallel argument tells against its algebraic premise. The straight rule, in this or in other forms, may indeed by open to criticism, but it cannot be refuted in this manner.

It follows that we may question Sir Karl Popper's description of this as 'an important and indeed brilliant discovery'; and Popper's generalisations of Miller's procedure are equally open to criticism. For example, the following albebraic argument is parallel to Popper's argument in II (the steps being numbered to correspond with his, except that his (iv) and (v) both correspond to my (iv), and his (vi) incorporates steps corresponding to both my (v) and my (vi)).

(i) $V(a/a = r) = r$
—the 'apparently intuitively satisfactory' assumption which is to be refuted.

(ii) $V(a/a = s) = s$
—from (i) by analogy.

(iii) For all x, if $s = 2r\text{-}x$, then $x = r$ iff $x = s$
—by algebra.

(iv) If $s = 2r\text{-}a$, then $a = r$ iff $a = s$
—from (iii).

(v) If $s = 2r\text{-}a$, then $V(a/a = r) = V(a/a = s)$
—from (iv), together with the fact that if $B \equiv C$, then the value of a, given B, is equal to the value of a, given C.

(vi) If $s = 2r\text{-}a$, then $r = V(a/a = r) = V(a/a = s) = s$
—from (i), (ii), and (v).

(vii) If $s = 2r\text{-}a$, then $r = s$
—from (vi).

But (vii) is incorrect. Also, we can satisfy the condition of (vii) by substituting '$2r\text{-}a$' for 's' in (vii) ; this leads to :

(viii) $r = 2r\text{-}a$

and hence $a = r$, which is incorrect, since a and r are free variables.

If this argument were sound, it would compel us to reject (i) ; but (i) is correct, and this argument is unsound, and so is the argument of Popper to which it corresponds. The fallacy occurs in each case at step (vii), and is of exactly the same character as the fallacy in Miller's argument, though the additional complications wrap it up more effectively.

To obtain a contradiction, we must read (vii) as implicitly *triply* quantified, 'for all values of r, for all values of s, for all values of a' (and, correspondingly, we must read (viii) as doubly quantified, 'for all values of r, for all values of a'). *But we cannot read (vi), from which (vii) is derived, in this way.* Given the antecedent '$s = 2r\text{-}a$', '$V(a/a = s) = s$' is equivalent to $V(a/a = 2r\text{-}a) = 2r\text{-}a$', and in the latter equation we cannot take the 'r' and the 'a' as *independently* universally quantified. It would be nonsense to say here 'for all values of r, *for all values of* a, $V(a/a = 2r\text{-}a) = 2r\text{-}a$' ; once a value has been given to r, no separate variation of a is allowable within this statement. Thus (vi), by its very formulation, restricts a to a value equal to that of r ; but (vii), by leaving out the terms containing 'V', drops this restriction, and is therefore *invalidly* derived from (vi). (If this restriction were inserted into (vii) and (viii), the contradiction would, of course, disappear.)

An exactly analogous criticism applies to Popper's argument in II ; his 's' in (vi) stands for '$2r\text{-}fr(a)$', and in (vi) the 'r' and the '$fr(a)$' are *not* independently variable, while in (vii) and (viii) they are treated as if they were. This comes out still more clearly in the version of the argument given in III ; in equation (ii′), r and $fr(a)$ are not independently variable, since this equation means 'The probability of A, relative to the observation that the frequency of a is equal to $2r$ minus the frequency of a, is $2r$ minus the frequency of a' ; but (iv′), which is supposed to be derived from (ii′), is inconsistent only because r and $fr(a)$ are treated as being independently variable.

Again, the version of the argument which Popper gives in IV turns upon the use of the formula '$E^{M}_{s_N}, a$' which, as Popper explains under (a) describes evidence which includes the observation that $s_M = s_N$, that is, that the number of observed individuals with the property M is equal to the number that lack this property. Consequently equation (3), namely

$$c\left(H^{M}, E^{M}_{s_N}, 3\right) = s_N/(s_N + 3)$$

holds for the *particular* value of s_N assigned by the evidence mentioned, namely $s_N = 3$; equation (3) is *not* implicitly quantified 'for all values of s_N', but the results obtained in (5) and (6) are absurd only if these equations are taken as so quantified ; if they are taken thus, then they are *invalidly* derived from (3), since it is not and cannot be so quantified.

There is one further point. Sir Karl Popper, in his Comment, argues that the contradiction attaches not to the 'straight rule' as such, (Rc) or (Rcr), but to a form of this which 'makes *essential use of the probability calculus*', the reason being that this calculus contains the rule (Z′), If $B \equiv C$ then $P(A, B) = P(A, C)$. Now it is true that this rule is essential for the derivation of the contradiction : it is used in the proof of step 4 in my formulation of Miller's argument and in the proof of step (vi) of Popper's argument in II, and so on. But it is not this rule that is the source of the contradiction, or even the combination of this rule with the special form of the straight rule used. For in my algebraic parallel, the analogue of (Z′) holds, namely 'If $B \equiv C$, then the value of a, given B, is equal to the value of a, given C', and so does the analogue of the special form of the straight rule, '$V(a/a = r) = r$. But no contradiction would follow from them without the invalid moves made at steps 6 and (vii). Similarly, without the fallacies introduced at the corresponding steps in the arguments proposed by Miller and Popper, no contradiction would follow from the combination of (Z′) with the special form of the straight rule ; there is, consequently, no need for the 'one hopeful way out' mentioned at the end of Popper's Comment.

J. L. MACKIE

ON A SO-CALLED SO-CALLED PARADOX: A REPLY TO PROFESSOR J. L. MACKIE

IN section (I) of this note[1] I hope to show that Professor Mackie, in his elaborate comments, nowhere comes near to criticising either the argument I gave in 'A Paradox of Information', or the arguments of Sir Karl Popper's note. In section (II) I shall briefly explain some of the ideas that may have, it seems, misled Professor Mackie.

I

Whether our arguments (Professor Popper's and mine) are valid or invalid need hardly be argued here. For although they were set out with some care, so that each step could easily be identified and referred to, Professor Mackie does not pin-point in them one invalid step. Instead he says (in the fourth paragraph of his note) 'let us state his [Miller's] proof a little more fully', and then proceeds to state, very fully indeed, two arguments (numbered stepwise 1-8) which may or may not be invalid (I am prepared to grant for the sake of argument that they are invalid), but which, if invalid, certainly do not share their invalid steps with any of our arguments.

This is quite obvious, for the invalid step (5 to 6) pointed out by Professor Mackie consists in the universal quantification of his variable 'a'. But no such quantification is involved, explicitly or implicitly, in our arguments where, in fact, 'a' is *not a*

---

[1] I wish to thank Professor Popper for his assistance, and also Allan Findlay for checking all the arguments.

variable but a constant : the name of the kind of event—*5-uppermost with this die.* (This was chosen merely to make it more obvious that the result of *our* derivations, namely step 5, is absurd.) Nowhere does either of us attempt to treat this constant as a free variable; nor do we try to put it into a universal operator.

If, however, 'a' is treated as a metalinguistic variable (rather than as a constant), it can be put validly into a universal operator in a step immediately following 1. So we can replace 1 by

1′. For all a, for all r,

$$P(a \text{ at next throw/the frequency of } a\text{'s} = r) = r$$

and from this we can of course proceed by instantiating 'a' and 'r' independently of one another.[1]

II

Professor Mackie's second argument (like his almost identical third argument, which he numbers (i)-(viii)) seems to explain why he feels so certain that our arguments (especially here Professor Popper's) are invalid. He rightly points out that a similar contradiction can be derived by a 'parallel argument' from what he believes to be 'an undeniable algebraic statement, namely "For all x, the value of a, given that $a = x, = x$" ', which he also writes

1. $$\text{For all } x,\ V(a/a = x) = x.$$

But though this statement looks convincing enough, it is neither 'algebraic' nor true. Professor Mackie himself derives from it

5. $$1—a = 1/2$$

and, later (by *his* quantification which he takes to be responsible for the invalidity of *our* derivations),

$$1/2 = 5/6$$

He defends 1 on the ground that his derivation is fallacious, as perhaps it is. The steps down to 5 Professor Mackie does not contest, for he does not notice that 5 is sufficient to refute 1. The steps after 5 are merely 'additional complications' of Professor Mackie's.

What we have still to clear up is the apparently undeniable character of 1. Professor Popper has already explained in his note why statements of the form of 1 seem undeniable, and it is done with such clarity that I recommend that his note be

---

[1] Incidentally, this is just what Professor Mackie seems to deny. For, if I understand him correctly, he believes that when we instantiate 'r' we automatically confine the frequency of a's to the value given by the evidence statement ; that is, he believes that the evidence statement must be true. But this means that whenever he talks about 1 (or 1′) he in fact has in mind

(Rcr) For all a and r such that the frequency of a's $= r$,

$$P(a \text{ at next throw/the frequency of } a\text{'s} = r) = r$$

which, as Professor Popper has shown, is not at all the same as 1 and which, not surprisingly therefore, fails to yield a contradiction.

I should like to emphasise that the main point of my note is that Professor Mackie has altogether failed to criticise our arguments. It has, I hope, been established in the text, and this footnote is intended only as a further clarification of Professor Mackie's position.

read again with care. What has happened is that 1 has throughout been mistaken by Professor Mackie for

(Rc) For all x, if $a = x$ then $V(a) = x$

which is both undeniably true and undeniably algebraic. The difference between 1 and (Rc) is hardly noticeable in English, and this explains Professor Mackie's intuitive conviction that 1 is true. But it is very noticeable if we compare the admissible logical procedures in each case. (Rc) allows the application of *modus ponens* or the *rule of detachment* for hypotheticals, while 1 does not : instead it allow the derivation of an *equation* whenever the evidence statements are *logically equivalent.*

Though some algebraists might make use of hypotheticals such as (Rc), statements of the logical form of 1 have not so far been used in algebra (though they have been used by some probability theorists) ; it seems with good reason.

Professor Mackie says the fallacy in Professor Popper's argument ' is of exactly the same character as the fallacy in Miller's argument '. I accept this of course since there are no fallacies. But when he goes on to say that ' the additional complications wrap it up more effectively ' then I can only assert that the ' additional complications ' have been produced neither by Professor Popper nor by me. Moreover the resulting fog seems to have now become so dense that, although Professor Mackie has in fact himself taken the ' one hopeful way out ' which Professor Popper has described (and since refuted), he has failed to notice that he has done so, for he proclaims at the same time that there is ' no need ' for such a way out.

University of London DAVID MILLER

ORIGIN AND CONCEPT OF RELATIVITY (PARTS I AND II): REPLY TO PROFESSOR DINGLE AND MR LEVINSON

I AM grateful to Professors Herbert Dingle,[1] Max Born[2] and Paul Feyerabend[3] for appreciation of my paper[4] (Parts I and II).

I have shared with Dingle for the last few years his intense desire to grasp the scheme of things called relativity and to remould it nearer to the heart's desire. With this identity, it is even more obligatory for me to comment on certain aspects of Dingle's note.

(i) I cannot agree with Dingle that the question of priority does not arise because the three protagonists meant different things. They did. That is why I devoted a whole section of my paper to the question : Did the three—Poincaré, Lorentz and Einstein—mean the same thing in their ideas on relativity? But they meant much that was common. And that much mattered the most. To Poincaré and Einstein, relativity meant that relative but not absolute motion, could possibly be ascertained physically. Both regarded Lorentz transformation equations (L.T.E.) as the mathematical embodiment of the doctrine. Lorentz too posited L.T.E. to explain the null result of experiments undertaken to detect the absolute motion of the earth. Again, all the three postulated an upper limit of velocities equal to the velocity of light.[1]

---

[1] This *Journal*, 1965, 16 (63), 242

[2] Private communication. Born was kind enough to say that the appraisal was ' most satisfactory and just '.

[3] Private communication.

[4] *Origin and Concept of Relativity*, Parts I and II, this *Journal*, 1965, 15 (60), 286; 1965, 16 (61), 19. Part III appeared in the February 1966 number.

True, Poincaré (sometimes) and Lorentz suggested or implied that experiments to detect absolute motion are conceivable but they did not describe one, even in theory, at any rate till 1912[2] (year of Poincaré's death).

On the other hand, Einstein himself predicted 'absolute' effects when he concluded in his original paper that a 'moving' clock goes slower than a 'stationary' clock, violating his own dictum that 'the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest'[3] (rest or motion). Two violations of the principle of constancy of velocity of light by Einstein in his original paper were pointed out in my paper. One of these is the subject of Levinson's [4] note.

So, who meant what and how surely?

Undoubtedly, the central ideas (of relativity theories) were sufficiently unique, profound and sharp, and now there is evidence enough, for one to assert justly that Poincaré postulated the principle of relativity, which Einstein learnt and adopted, using Lorentz's mathematical formulation. Let us not chop logic and perpetuate a grave injustice.

(ii) As to Poincaré's belief in the ether, I would say that I was quite aware of it. I myself said that Poincaré up to the end offered lip-service to the ether. The passage in *Science and Hypothesis*, which Dingle does not concede me as showing that Poincaré denied the ether, runs as follows[5] (excerpts) :

> 'Does our ether really exist? . . .
> Many experiments have been made on the influence of the motion of the earth. The results have always been negative . . . according to the current theories, the compensation would be approximate and we might expect to find accurate methods giving positive results. I think that such a hope is illusory. . . . I do not believe, in spite of Lorentz, that more exact observations will ever make evident anything else but relative displacements of material bodies. . . . He (Lorentz) showed that the terms of the first order should cancel but not the terms of the second order. Then more exact experiments were made which were also negative. . . . No, the same explanation must be found for the two cases, and everything tends to show that this explanation would serve equally well for terms of the higher order. . . .'

We note : experiments involving second order terms (v^2/c^2) were mentioned. This passage left the impression on my mind that in it, Poincaré was denying the ether, not affirming it. But there is a sentence in the same passage (not reproduced above but quoted in my paper) which suggests that it is not absurd to assert that absolute motion can be determined but this suspended state of belief was quite in keeping with the evidence available to Poincairé—and now!

---

[1] Lorentz also postulated this limit in 1904. *The Principle of Relativity*, Dover, p. 13. Here, I do not repeat the references given in my paper quoted above.

[2] Later, Lorentz referred to the clock-paradox and asymmetrical ageing but still did not affirm that they supported his ideas more than Einstein's. *Das Relativitatsprinzip*, 3 Haarlemer Vorlesungen, 1914, pp. 31 and 37. *Problems of Modern Physics*, Ginn & Co., 1927, p. 254. Commenting on the latter, Larmor, then alive, wryly remarked, '. . . the theory is driven to conclude that a celestial journey lengthens life'. *Proc. Roy. Soc. Edin.* 1926–27, 47, 307.

[3] *The Principle of Relativity*, Dover, pp. 49 and 37 respectively. Faithful to relativity, Dingle has called Einstein's statement about the moving clock as a 'regrettable error'.

[4] This *Journal*, 1965, 16 (63), 246

[5] *Science and Hypothesis*, Dover, pp. 169–172, Incidentally, the correct date of first publication is 1902 and not 1903 mentioned by Dingle.

(iii) The above passage shows that at the time of writing of *Science and Hypothesis*, Poincaré did not mean by the term relativity, simply what is implied by Newton's first law, as Dingle says. Poincaré meant more. The second order terms arose in the explanation of Michelson-Morley experiment, and in 1902,[1] Poincaré was acutely seized of the result of this experiment and the related situation in electrodynamics.

(iv) But Dingle is right in pointing out with anguish, that relativistic and ether-theoretic ideas are used indiscriminately, as if there were no difference between the two. For example, in explaining the experiment of Wood, Tomlinson and Essen on the frequency of a vibrating rod when oriented differently, Whittaker[2] wrote, 'According to relativity theory, however, there should be a complete compensation of length, by a modification of the elasticity of rod according to its orientation'. The experimenters[3] themselves gave the same explanation, apparently on the advice of Eddington! According to relativity, no change in length, density, mass, elasticity or frequency will occur or be observed for a rod stationary within an inertial laboratory, no matter how the rod is oriented. On ether-theoretic considerations, the vibrating rod essentially functions as a clock, the rhythm of which is not dependent on orientation within a system although it is dependent on the absolute velocity of the system. However, all the clocks are equally affected within a system. Again, while discussing the clock-paradox, W. H. McCrea[4] wrote, 'The fundamental point is that it is not meaningful to speak merely of the motion of one body relative to another with no reference to the rest of the universe'. This is absolutely correct, but it is not relativity then, because we postulate 'properties corresponding to the idea of absolute rest'.

It is sometimes said that there is no *observable* difference between the relativistic and ether-theoretic points of view.[5] This is incorrect. As I explained in Part III of my paper, the time dilation factors for systems S′ and S″ are $(1-w^2/c^2)^{-\frac{1}{2}}$ and $[1-(v+w)^2/(1+vw/c^2)^2c^2]^{-\frac{1}{2}}$ where v is the velocity of S″ relative to S′ and w is the velocity of S′ relative to the system of the stars (ether), w being assumed to be in the direction of v. The *relative* dilation factor is then $[(1-v^2/c^2)/(1+vw/c^2)^2]^{-\frac{1}{2}}$ and not $(1-v^2/c^2)^{-\frac{1}{2}}$. (Similar considerations would apply to the Doppler-effect and mass-variation formulae). Also in such a case the clocks of S″ will be slower than those of S′ and not vice versa.

(v) That brings me to the principal point of Dingle's. May I reformulate it as follows? What exactly is the principle of relativity?[6] Is it valid? I have dealt with these questions in part III of my paper. Briefly, the situation is as follows. Relativity means: relative but not absolute motion is physically ascertainable. This implies that all inertial systems are equivalent:

---

[1] In fact even earlier, See *Rapports du Congrès de physique de* 1900, Paris, 1, 22-23, and a remark by Lorentz in *The Principle of Relativity*, Dover, p. 13

[2] *History of the Theories of Aether and Electricity* (1900–1926), 1953, p. 43

[3] *Proc. Roy. Soc.* (A), 1937, 158, 607–608 , [4] *Discovery*, April 1957, p. 175

[5] I quoted V. Laue, and Margenau and Mould in Part II of my paper. W. Rindler, similarly, says that the differences are only philosophical: *Special Relativity*, Oliver and Boyd, 1960, p. 21. I was myself not altogether clear on this point in Part II of my paper.

[6] The earlier one, assumed for inertial systems. Einstein extended this to all systems having arbitrary motions. V. Fock considers the principle valid for inertial systems and those in which 'corresponding' physical processes occur (masses may then be present), but not for all systems in general. *The Theory of Space Time and Gravitation*, Pergamon, 1964, pp. 178-182, 369-375. According to the ideas presented here, only relativity of

(*a*) Internally, i.e. for the laws of physics *within* respective systems.

(*b*) Externally, i.e. when observations involving more than one system are made, effects dependent on the relative velocity but not on any absolute velocity can ever be detected.

The mathematical expression of the theory is said to be epitomised in Lorentz transformation equations (L.T.E.), but this is not quite correct. L.T.E. can be derived from the postulate of the constancy of the velocity of light, *alone*, without appeal to the principle of relativity. However, L.T.E. imply that:

(*a*) The laws of physics are internally the same *within* all inertial systems, since $\beta = (1 - w^2/c^2)^{-\frac{1}{2}}$ and space and time measures remain *constant* for each system. (The metric within each system is then reducible to the form: diag. $[+1, -1, -1, -1]$).

(*b*) No experiment using electromagnetic propagation could reveal absolute motion. This follows directly from the constancy of light-velocity.

Thus L.T.E. indicate a high degree of relativity but not complete relativity. An analysis of L.T.E. shows that time (and space) measures of different systems are different. This compels one to recognize the system, assumed at rest in deriving L.T.E., to be the preferential system of the stars (external galaxies). Indeed, absolute effects observed with accelerated motion demand this absolute system. The relevant experiments made so far have not been sufficiently accurate to determine the absolute velocity of the earth but this is possible.[1]

(vi) The last point of Dingle's: The velocity of light (c) compounded with any velocity $v < c$ is c in *magnitude*, but the direction is different by the angle of aberration. This follows from L.T.E.[2]

H. B. Levinson[3] has exhumed two old discussions of mine which undoubtedly contained a mistake. In extenuation may I say, what Levinson grants, that the mistake was the same as the one Einstein made in his original paper and no one criticised earlier?

In the filter experiment, two green spots will be seen in both the frames, separated by the angle of aberration. Here too the situation is symmetrical and satisfactory for relativity.

However, the essential arguments and objections of these discussions are still maintained in part III of my paper, Origin and Concept of Relativity.

Room No. 28A
Baroda House Annexe 1
Curzon Road, New Delhi, India

G. H. KESWANI

space, i.e. equivalence of all points in space, exists. Even this is only approximate, and fluctuations in the distribution of masses around, and in their motions, could affect the local laws of physics.

[1] We may recall H. Bondi's remark, '. . . it should not cause too much surprise if a locally observable consequence of this cosmically defined absolute velocity were discovered one day', *Report on Progress in Physics*, 1959, **12**, 97. Also see *Proc. Roy. Soc.* (*A*), 1962, **270**, 311.

[2] W. Pauli, *The Theory of Relativity*, Pergamon, 1958, pp. 16–17.

[3] This *Journal*, 1965, **16** (63), 246

REVIEWS

The Retreat to Commitment. By William W. Bartley III.
Chatto & Windus, London, 1964. Pp. xii+223+iv. 30*s*.

CHAPTERS iv and v of this book are good, and are worth serious consideration. The remainder of the book consists, firstly, of a journalistic survey of what the author regards as the current situation in Protestant thought, particularly in America, and then of an application of the central arguments to this situation. The reader is bombarded with italicised words and phrases, and irritating claims to originality which may dissuade him from attempting the important central section of the book. It is this section which makes such stylistic criticism of minor moment.

Bartley sets out to examine the question whether 'the choice between ultimately competing religious, moral, and philosophical positions is, in the last resort, *arbitrary*. For example, is an individual's decision to become a *rationalist*—even from a rationalist point of view—any less subjective, relative, arbitrary, irrational than an individual's decision to become a Christian?' He refers to this problem 'alternatively as the dilemma of ultimate commitment and as the problem of the limits of rationality'. Bartley regards the 'only serious argument for Christian commitment today' as the *tu quoque* argument, which might be described in this way: in order to stem an infinite regress arising from the reapplication of the question 'How do you know?' rationality, itself, must be accepted as limited to the extent that everyone must (logically) make *one* dogmatic, irrational, unjustified and unjustifiable assertion, which may be variously called his ultimate presupposition, the first premise etc. The rationalist and the Christian are thus, ultimately, in the same position: but one who makes such an irrational commitment is in no position to criticise another who makes a different commitment. Bartley aims to show that it is possible to choose in a non-arbitrary way among competing, mutually exclusive theories, and regards this aim of defeating the *tu quoque* as 'the fundamental problem of modern philosophy'.

Accepting Popper's argument that the western philosophical tradition is authoritarian in structure, and that questions of the form 'How do you know?' invite authoritarian answers—the bible, the leader, the intellect, the sense-experiences—Bartley argues that such questions should be abandoned. There are separable reasons for this policy. Firstly, 'all proposed intellectual authorities have turned out to be both intrinsically fallible and epistemologically insufficient'. Next, the view that *only* positions which can be justified or established by rational argument are to be accepted, cannot itself be justified by appeal to rational criteria, and if true must be rejected. Any view that a rationalist is *committed* to rationalism is an inescapably fatal concession to irrationalism. Lastly, and above all, there has been a repeated fusion of the ideas of criticism and justification, and any continual demand for justification logically issues in commitment (i.e. irrationalism), because demands for justification *are* satisfied by appeals to authority.

The position to be adopted is this: 'since there are no guarantees or criteria of truth, no ways of definitely deciding, I can never know for sure whether what I believe to be true is in fact so'; but the aim must be to expose all our beliefs, conjectures etc., to maximum criticism in order to eliminate as much error as possible; even this basic philosophical position itself is open to criticism. (Ayer, of course, had remarked in the first chapter of *Language, Truth and Logic*: 'A hypothesis cannot be conclusively confuted any more than it can be conclusively verified.') All justification is to be abandoned, for a position may be held rationally without needing justification—provided it is held open to and survives severe testing. This kind of 'rationalism' could be refuted, for example, by showing that some of the unjustified standards necessarily used in this (or any) position were *uncriticisable*, and had to be accepted as such in order to avoid circularity. By this recommended move, Bartley argues, one may avoid the *tu quoque* position of accusing a man of protecting some belief from criticism by irrational commitment to it.

Bartley emphasises that his new style rationalist is of course *convinced* of the truth of his position, although not committed to it. He also argues that he is not simply offering a redefinition of 'rationalist'. His aim has not been to explicate a meaning but to remove 'an ordinary assumption about rationalism which prevented the solution of the problem rationalism was intended to solve'. He considers, next, what is to count as criticism. Here the claim is reiterated that we can neither decisively establish *nor* refute any theory, 'for any theory will be refuted only relative to our acceptance of critical arguments that are incompatible with it'. Four means of eliminating error are mentioned: is the theory in question consistent, is it empirically refutable, is it in conflict with any scientific hypothesis, does it successfully solve the problem it was intended to solve? Bartley also discusses how his position differs from Quine's 'Two Dogmas of Empiricism'. Basically it is over the question whether in principle logic is revisable. Bartley argues that 'we cannot regard logic as part of the set of beliefs that are put to the test in critical discussion, for the notion of testing and revising in accordance with results of the test presupposes logic'. This is not to say he is *committed* to logic; to reject logic is to reject argument, of which it is a presupposition. *Not* committed (171); and yet: 'I have stated an absolute presupposition of argument to which we are committed not *as human beings*, because of our biology, psychology, or sociology, but as *arguers about the world*' (172).

There remains room for only a few queries. What is the status of the claim 'we can never know'—not only do we need to know *how* we know this, but also what counts as evidence and how meaning is given to the expression. Is it a factual lament, a deduction from premisses (themselves unquestioned), or what? It is not entirely clear how the recommended methodology stands in respect to the variously interpretable concepts of meaning, truth and evidence. Bartley rightly refers us elsewhere for elucidation of some of his claims; nevertheless he writes: 'we can decide freely and irrationally, as a matter of taste, only between two theories against which there exist no criticisms one is unable to defeat.' Presumably such subjects as morals and aesthetics fall under this rubric; or perhaps they are already demarcated as 'unscientific' and not open to the falsifiability criterion. Bartley would clearly endorse Oscar Wilde's remark that 'an idea that is not dangerous is unworthy of being called an idea at all'. He may be surprised to find this additional remark of Wilde: 'It is criticism that, recognising no position as final, and refusing to bind itself by the shallow shibboleths of any sect of school, creates that serene philosophic temper which loves truth for its own sake, and loves it not the less because it knows it to be unattainable.'

Peter Jones
Edinburgh University

Philosophy of Science: The Delaware Seminar. Vol. 2, 1962-1963. Edited by Bernard Baumrin.
John Wiley & Sons, London, 1963. Pp. xviii+551. 109*s.*

The aim of the second volume of this series is to examine most of the basic philosophical questions concerning the physical sciences.

The first part, 'Scientific Explanation, Prediction and Theories' contains five papers, all of which purport to attack either the whole or some part of Hempel's account of the logic of explanation. Feyerabend reproduces his, by now, well-known criticisms of modern empiricists who are, he claims, the new perpetrators of the sins of the 'school-philosophers'. His purpose is to show that their account of scientific theories leads to dogmatism and the repression of scientific advance. The attack leans heavily on the refutation of two conditions which are said to be fundamental to modern empiricism. These are (I) the Consistency condition: 'Only such theories are admissible in a given domain which either *contain* the theories already used in this domain or which are at least *consistent* with them inside the domain' (p. 10), and (II), the condition of meaning invariance: 'Meanings will have to be invariant with respect to scientific progress; that is all future theories will have to be phrased in such a manner that their use in explanation does not effect what is said by the theories, or factual reports to be explained' (p. 10). (I) is derived from Hempel and Oppenheim's paper and (II) from Nagel's account of reduction. These conditions are said not be invariably accepted by science and to be inherently unreasonable. However, to some extent I think Feyerabend has set up a straw man. (I) is said to be a more explicit statement of (A): ' the consequence of a satisfactory explanans, T, inside D [the domain], must be compatible with the explanandum, T' (p. 9), if we take for granted that more general theories are always introduced to explain existing successful ones. It is not obvious that empiricists do take this for granted, but even if this is the only reason for introducing a new theory, (I) still is not ' a more explicit formulation ' of (A). For (A) gives a relationship between an *explanation* and what is explained, whereas (I) gives a condition for the *admission* of a new theory. It is not difficult for Feyerabend to show that (I) leads to absurd consequences, and hence that it is an absurd condition for the admittance of a theory into one science. But none of his arguments presented here show that (A) is an absurd condition for explanation.

(II) is more interesting, but even here there is a slide from a condition for successful reduction to a condition for the correct interpretation of any theory in a given domain. In effect, Feyerabend attacks the distinction between observation and theoretical terms, but although it is a mistake to suppose that no terms change their meanings during the course of scientific development, his argument that the meanings of *all* the terms in a theory are dependent on that theory leads to a paradoxical situation when taken together with his positive thesis. For he maintains that criticism is only possible if we work with alternative theories within a given domain. But for theories to be alternatives, they must proffer explanations of the same facts. In other words, at least part of the meanings of the terms used to describe the phenomena to be explained must be independent of the various competing theories. This is one version of the source of (II) and to reject this makes it impossible to approach science critically. What one can say is that the meaning of a ' prediction-sentence ' derived from a theory is dependent on that theory, but this sentence is related to some sentence used to describe the phenomenon, whose meaning is independent of all the competing theories in the domain. But to say this is only to recognise the problem, not solve it.

Sellars's arguments, in ' Theoretical Explanation ' for a realist interpretation of theoretical entities, depend upon the distinction which Feyerabend rejects. Using it, he is able to show that at least some correspondence rules, relating theoretical predicates to empirical predicates, do make use of the notion of identity, which may either be ' taken seriously ' or treated as syntactical statements in the manner of the instrumentalists. Since they are, in fact, taken seriously by scientists, he asks why philosophers should refuse to do so.

The alleged symmetry between explanation and prediction has often been attacked and Rescher provides further reasons for rejecting it. His argument rests on the possibility of finding some discrete-state systems whose states can be ' explained ' by means of a statistical theory, but cannot be predicted. Instead of rejecting such theories as explanations, we should admit them and recognise that the aim of science is not to explain (in the Deductivists' sense) but to 'rationalise' or 'systematise'. (These last two terms are left unexplained.) Now Rescher does not show that it is *impossible* to provide an explanation of the states of the system by reference to some underlying mechanism (yet to be discovered) which would enable us to predict these states. So, if the *pragmatic* justification of explanatory theories for which this symmetry holds is that they provide us with a greater possibility of control of our environment, then we should not stop searching for them because the statistical theories we have are rightly called 'explanations' or because explaining is a different activity from predicting. One can 'systematise' for more than one reason.

Dr Bromberger suggests that several of the problems raised by the Hypothetico-Deductive (H-D) account of theories are of little relevance to science and that there are alternatives which have its virtues but not, by implication, its vices. The old account, he claims, does not distinguish between two types of theory, that which can be accepted or rejected, theory$_2$ (The Kinetic Theory of Heat) and that which cannot be accepted or rejected theory$_1$ (The Theory of Heat). He develops a formal account of the relation between theory$_1$ and theory$_2$, which turns out to be not so much an alternative to the H-D account, but an elaboration of it.

For some time attention has been given to problems raised by inaccuracy in scientific methods and theory. Scriven examines some of these, but in a limited way. He tends to dismiss the inaccuracy of data-statements as not very important, but his treatment of the inaccuracy of laws is more detailed. His aim is to show that there can be no formal account of scientific theories and his main argument is that because the data of a theory and the incorporated laws are inaccurate (in different ways), the pattern of inference in the theory cannot be formalised since the inaccuracies cannot be made explicit. No argument is given to show *why* they cannot be made explicit, nor that statements about the implicit inaccuracy of laws are statements within any particular theory rather than statements about the interpretation of theories in general. And there is no reason why 'completeness claims' (e.g. 'there are no other significant forces acting') which he says is one feature of the inaccuracy of laws, cannot be made explicit in a theory. The principal difficulty which faces those who want to give formal accounts of theories is how to characterise the relation between a law-statement which is known to be an approximation and data-statements (which may also be approximations) so as to preseve the deductive structure of theories. Pointing at the problem does not show that it is insoluble.

All five papers raise doubts about the traditional H-D account of theories, but I do not think Feyerabend, Bromberger and Scriven have successfully shown that the philosophical basis of the account, or the account itself is untenable in any modified form.

In Part II (' Space and Time '), Shapere discusses the relation between the use of terms by scientists and ' ordinary ' people, taking as his examples ' space ' and ' time '. Puzzles such as the paradoxical nature of ' space is expanding ' have led to various accounts of the relationship which Shapere dismisses. The only way to solve such puzzles is ' to trace the relationships between ordinary and scientific usages of the terms involved in the paradoxes, and to examine the reasons why the departure of usage have been made ' (p. 155). From this position he goes on to argue that, since we want to understand the shifts of meaning of these terms, the history of science and philosophy of science are *logically* related. Against this, it might be objected that, although such a historical enquiry may resolve the paradoxes for the ' ordinary ' language-user, the philosopher is concerned with the analysis of *concepts* of space and time as found in the sciences and elsewhere. This will be a prior activity to the historical one, and the latter will be impossible without the former. In other words, if we want to understand shifts of meaning of this type, philosophical analysis precedes historical enquiry. Grünbaum in a savage attack on Polyani and Whittaker, admirably demonstrates how a prior understanding of the philosophical conceptions of relativity theory is necessary for an account of its genesis. Also, those who are tempted to reconcile incompatible accounts in the history of science by appeal to an analogy with quantum mechanics would benefit from Grünbaum's paper.

Putnam examines Grünbaum's own philosophy of Geometry and it is a pity that the volume contains no reply from him. Hill and Suppes, in Part III, write on philosophical problems in quantum mechanics, while Pais, in an entertaining paper, discusses scientific and human ones. Part IV contains papers on Induction and Measurement. Salmon gives a simpler account of his vindication of induction, based on earlier papers. Ellis contributes an elegant paper on ' Derived Measurement, Universal Constants and the Expression of Numerical Laws '. He first shows that there is a convention governing the expression of such laws, that we express them either with respect to particular scales or with respect to particular classes of scales. From this argument and his initial discussion of scales he then shows that there is a convention for the expression of derived measurement too. We should always ' use similarly defined derivative scales for the measurement of all other quantities [not measured directly] ' (p. 387). One consequence of these conventions is the *invariance* of relative magnitude with changes of scale. Hence, Bridgman's postulate of the absolute significance of relative magnitude is placed on a firmer foundation.

The remaining papers are a mixture of philosophy, speculation and popular accounts of science. Hanson attacks the confusion in contemporary cosmological theories, while Wheeler presents his own theory and discusses some of its implications, scientific and human. In accordance with what appears to be now a well-established tradition in collections of this kind, there is a paper on the relation between physical and biological science. Pollard, in effect, gives a popular account of recent developments in molecular biology. Lindsay discusses the 'Thermodynamic Imperative' and its relation to Ethics. Sterling McMurrin writes on problems in contemporary education.

The title of this book and its statement of aim is misleading. There is virtually no interchange of ideas between contributors, which would have been particularly welcome in Part I, and its breadth is so wide that less than three hundred pages out of five hundred and fifty are devoted to philosophical problems which arise specifically in the physical sciences.

DAVID HIRSCHMANN
Bristol University

Forces and Fields. The Concept of Action at a Distance in the History of Physics.
By Mary B. Hesse.
Thomas Nelson & Sons, London, 1961. Pp. x+318. 35*s*.

DR HESSE has written an excellent and fascinating book which should be read and studied closely by all philosophers and historians of science. It is, as the sub-title informs us, 'a study of action at a distance in the history of physics'. But it contains far more than this; it is, as well, a model of that history which Lord Bolingbroke warmly recommended—'Philosophy teaching by example'—and a penetrating analysis of such fundamental concepts as mechanism, analogy, model, atom, and force.

At the beginning, Dr Hesse gives us her framework in a chapter on the logical status of theories. Historians would do well to ponder this for most histories of science keep philosophy at arm's length, if it is even allowed to enter (overtly) at all. By explicitly showing how various philosophical aspects will necessarily affect historical interpretations, and by permitting the reader to know her philosophical position, Dr Hesse has provided a solid base for what follows. She has also gone deeper than any other work of a similar kind.

Following her philosophical preamble is the history which reaches from antiquity to modern physics. Along the way, much of the history of physics is given new meaning and illuminated.

There seems to me to be only one major aspect which deserves criticism and this follows necessarily from Dr Hesse's point of view. She is interested primarily in the *logical* development of the concepts involved. To be sure, there are minor expeditions into anthropology and mythology in the early chapters, but the historical line becomes increasingly leaner and tauter as the present is approached. There is nothing wrong with this, but it does serve to obscure some historical factors which ought to be rescued. I must here, with apologies, turn to my one area of strength—Michael Faraday—to illustrate this. Dr Hesse gives an excellent analysis of Faraday's ideas but she neither raises nor answers the question of *why* Faraday was led to reject action at a distance. It can be argued that the answer lies outside the realm of the logic of science and is to be sought, instead, in Faraday's personal life. This, in turn, means an examination of the historical milieu in which Faraday lived and some attention to the manuscript record.[1]

The same argument can, I think, be directed against the chapter on the theory of relativity. Dr Hesse here implies that the Michelson-Morley experiment is crucial; Einstein's 1905 paper then follows logically. But, there is considerable evidence to show that Einstein had never heard of the Michelson-Morley experiment and so, in order to understand *why* he was led to special relativity, we must abandon logic and pursue Einstein, the person, through his manuscript remains until an answer (if any) is found.

Unfortunately, what we do not know at this point is whether Faraday and Einstein are typical or whether they are the exceptions to the historical rule of scientific development. There has simply not been enough work done, in depth, on the biographical side of the history of science to judge whether one *must*, in all cases, delve into the personal lives of the creators of scientific theories, or whether the internal logic of science itself is a sufficient guide. My feeling is that non-scientific and intensely personal factors do play an important role in the development of science, but it is only a hunch, not a documented position. Dr Hesse, therefore, may well be right in ignoring the biographical dimension, but it should be pointed out that she may be vulnerable here. What she has given us, however, far outweighs what has been passed by.

The book is well printed and contains a good index.

L. PEARCE WILLIAMS
Cornell University

Form and Strategy in Science. Studies dedicated to Joseph Henry Woodger on the occasion of his Seventieth Birthday. Edited by John R. Grigg and F. T. C. Harris
D. Reidel Publishing Company, Dordrecht, Holland, 1964. Pp. vii+476. Df 1.38.

PROBABLY Joseph Woodger is the only person qualified to review this book, and to give an adequate idea of its contents even he would need more space than the editor could allow. Apart from a *curriculum vitae* of our ex-president, a list of his publications, and a foreword by Sir Cyril Hinshelwood, it comprises twenty-six papers. These are on subjects as diverse as formal semantics, probability theory, models of animal growth, and taxonomy; and they vary in size from three-and-a-quarter pages (in which W. V. Quine notes four not altogether creditable causes for the common belief that the simpler hypothesis stands the better chance of confirmation) to forty pages (in which Robert Rogers surveys formal semantics).

---

[1] See my *Michael Faraday, A Biography* (London, 1965) especially Chapter II where these factors are considered.

Eleven contributions are by biologists and concerned with special problems in the methodology of their own study. About these I am not competent to say more than they show the wide range of Woodger's influence among his professional colleagues, especially in the United States of America, where eight of the eleven authors teach. But I may perhaps add that the papers in this group which interested me especially were one by Paul G. 'Espinasse on Genetical Semantics and Evolutionary Theory and one by John Davidson on Animal Organization as a Problem in Cell Form. 'Espinasse suggests that an alteration of linguistic usage may be required to accommodate Waddington's discovery of genetic assimilation, whereby a character favoured in natural selection can change its genetic basis within a species over time. Davidson describes how the organs of polyploid amphibians may show the same general structure as those of haploid organisms of the same species although they are composed of fewer cells and sometimes therefore of cells with different shapes. In each of these papers we learn about elegant uses of experimental methods to investigate questions suggested by theoretical considerations.

Among non-biological contributions printed together under the title 'Philosophy of Science' those by E. W. Beth, Robert Rogers, John G. Kemeny, and R. M. Martin form a coherent group, being all alike concerned with the investigation of language by the methods of Carnap and Tarski. Beth's article on the Relationship between Formalized Languages and Natural Languages is a programmatic piece in favour of the thesis that 'in the logical analysis of natural language (as a means of expression for deductive theories) we ought to start from the insights obtained as a result of the construction and the study of formalized languages', and Rogers's work (mentioned above) is a clear summary of results up to date. Kemeny's paper, which is also partly historical, contains a defence of Analyticity versus Fuzziness and a plea for Carnap-style explication of concepts in contrast to the deliberately non-committal attitude of Quine or the reverential reporting of ordinary language by the followers of Wittgenstein. Martin's work is more ambitious, being a preliminary version of some of the material due to appear in his *Intension and Decision*. Together these papers provide a very useful introduction to the kind of philosophising which their authors commend.

In an article on Creative and Non-creative Definitions in the Calculus of Probability Sir Karl Popper discusses (with special reference to his own work on the axiomatisation of probability theory) the peculiarity of certain definitions whose presence within a deductive system makes possible the derivation of otherwise underivable theorems that contain neither the defined expression nor any expressions introduced by means of it. The creativeness of a definition is shown to be relative to the rest of the system in which it appears, and its peculiarity is said to be due to the fact that (like the 'real' definitions of older philosophy) it provides for the existence of something whose existence would not otherwise be deducible. In some systems the creativeness of a definition can be eliminated by addition of a special existence axiom, but only if the means of expression at out disposal in the system under consideration allow for the formulation of such an axiom, and this condition may not always be fulfilled. This paper seems to me to bring out in an interesting way the importance of distinction of type among variables. In the same section (on Logical Analysis of Theory Structure) there are two other interesting papers of technical logic. One by Frederic B. Fitch on the Algebraic Simplification of Redundant Sequential Circuits is an extension of a use of Boolean Algebra first suggested by Shannon. The other is a careful and impressive paper on Aristotle's Syllogistic and its Extensions in which Czeslaw Lejewski shows how successive enlargements of Aristotle's system can lead up to Lesniewski's ontology (which is not metaphysics but a theory of the copula.)

In the section called Models in Science I can mention only a discussion of Probability Models and Thought and Learning Processes by W. Mays, an argument for a Game-theoretical Approach to Organisation Theory from Olaf Helmer, and an account of an example of Design By Natural Selection in a computer from B. Dunham and his associates in the IBM Research Laboratory. The second of these seems to me to indicate a very promising new field of study.

WILLIAM KNEALE
Oxford University

The Explanation of Behaviour. By Charles Taylor.
Routledge and Kegan Paul, London, 1964. Pp.x+278. 40s.

Is behaviour behaviouristic? Mr Taylor has written an excellent book claiming that this an is empirical question and that the answer is—No. The behaviour of living things—in particular that of the higher animals and man—cannot be explained in terms of principles drawn from the non-biological sciences, in terms of laws linking separately identifiable physical events. Living organisms are different in nature from inanimate things, and must be described and explained in terms of special (and irreducible) categories. Behaviour must be described in terms of action and explained in terms of purpose.

This means more than that *if* behaviour is described in terms of action, *then* it must be explained in terms of purpose; and more than that we do, in fact, ordinarily describe behaviour in terms of action and must *therefore* accept explanations in terms of purpose. To Mr Taylor, ordinary language is no more than ordinary, and might well embody assumptions which will one day be shown to be mistaken. In particular, it might rightly draw a distinction, but wrongly assume a certain rationale for that distinction (cf. Aristotle on ' natural ' and ' violent ' motion). Reminding us of the arguments of Anscombe and Kenny, Mr Taylor shows that ordinary language assumes that behaviour is *ultimately* purposive, that no explanation in terms of purpose can be reduced to or replaced by a non-teleological principle. That purposive explanation cannot be reduced to a non-teleological form is a philosophical matter; that such explanation of behaviour cannot be replaced by any non-teleological form is an empirical matter. Since ordinary language assumes that no such replacement is possible, concepts such as *action, intention, desire* . . . would change radically or be discarded should scientists ever succeed in producing satisfactory non-teleological explanations of behaviour. However, Mr Taylor thinks it highly unlikely that we shall ever be faced with this dilemma; in the latter part of his book he argues convincingly, and in great detail, that peripheralist (S-R) psychological theories either fail to explain behaviour or do so merely in virtue of *ad hoc* hypotheses involving teleological concepts such as *goal, intention* or *purpose.*

Mr Taylor distinguishes between teleological and purposive explanation, the latter being a sub-class of the former and involving the notion of *intentionality*, or description under which the situation is seen by the agent. A teleological explanation involves laws in terms of which an event's occurring is held to be dependent on that event's being required for some end; moreover, the event is not only explained, but *identified* with reference to the future event or end, since the system is such that there is no finite list of intrinsic descriptions of those events (and those only) which occur in order that the end may be achieved. This is at odds with the atomism of traditional empiricism; but we assume too easily that scientific method must be defined within the limits of this particular epistemology, and if we question this assumption we question also the traditional objections to teleology. That a system is in such a state as to require a given event if a certain result is to accrue can be *perfectly observable*, and can be established independently of the evidence provided by the occurrence of the event itself. And if we give up our atomistic assumptions, we are no longer forced to think of *purpose* as an unknowable entity somehow exercising a mysterious backward causation, as the queerest of all psychological ' faculties '. Purposiveness is an observable characteristic of a system as a whole, and the behaviour of such a system must ultimately be explained in terms of (observable) inherent tendencies. Such tendencies provide *final* explanations of behaviour (cf. McDougall's hormic psychology, and Anscombe on ' desirability characteristics ') in their own right; though there is nothing to stop us asking for physiological explanations of the conditions necessary (but not sufficient) for these tendencies to have effect; and an evolutionist might point out that if a species did not have a certain general tendency, then it would not survive at all. No machine is—nor could one be—truly teleological, since it would be absurd to speak of its 'inherent tendencies' and to use these as final steps in the explanation of its behaviour. Still less could a machine be purposive, for this would require its having an 'intention' which could account for its 'action' in a non-contingent manner.

Mr Taylor's discussion of the peripheralist psychologies of theorists such as Hull, Nissen, Spence, Deutsch, and MacCorquodale and Meehl is masterly. Like McDougall he points out the crucial ambiguity of the term 'behaviour', he questions the existence of any S-R unit identifiable independently of the animal's goals, he focuses upon selective attention as a central behavioural characteristic which S-R theory cannot account for: and he shows that the importance of the phenomenon of latent learning is not, as is usually assumed, that it challenges hedonistic theories of motivation, but rather that it results in a *general* spatial orientation which can later be utilised to reach specific goals, which orientation cannot be accounted for in S-R terms. Over and over again he shows how various S-R psychologists have filled out their theories with hidden assumptions of purposiveness, with misleading neologisms which allow them to attribute intentionality to the animal's behaviour. To the extent to which they eschew such props, and avoid *ad hoc* hypotheses, their theories cannot explain the selective, goal-seeking activity of the animal, nor predict its behaviour in a open environment. Even ethologists such as Tinbergen, who try explicitly to account for the *directedness* of behaviour in terms of innate releasing stimuli, fail to explain the relatively plastic behaviour of animals high in the phylogenetic scale unless we allow their terms to have implicit teleological reference.

Of the many questions which remain with me after reading Mr Taylor's book, the outstanding one is whether 'centralist' psychological theories are open to the same criticisms as peripheralist ones. Much recent work has led to the postulation of internal models of the environment built up by the animal during learning, and used as intermediaries in his selection of behaviour when approaching his goals. Miller, Galanter and Pribram have stressed the notion of internal templates as standards for problem-solving, related to hierarchical systems of goals. Does this type of notion suggest a way in which intentionality may be more fruitfully dealt with by the psychologist? If an animal behaves in a certain way in the attempt to achieve some goal because he relies upon a certain model or template representing his environment, can we say that his behaviour (or his reaching the goal) is *contingently* connected with his reliance on this model? That an animal, after latent learning, chooses the *true path*, is surely non-contingently connected with his having achieved a general spatial orientation in his environment. The question is whether the psychologist, or the neurophysiologist, can give more specific content to such notions without destroying their teleological character.

A second outstanding problem is that of how we are to agree upon a classification, or even a listing, of the 'inherent tendencies' which will act as final steps in the explanation of behaviour; connected with this is the question of whether any motives exhibit the property of *functional autonomy* (cf. Allport). This is equivalent to asking what should be our rationale for identifying the basic psychological units of explanation, and this question Mr Taylor has not answered. But he has shown decisively that *stimulus* and *response* are pseudo-units, and that S-R 'correlations', insofar as they are explanatory at all, are based on a rationale which—being teleological—is theoretically denied by the very theorists who rely upon it.

MARGARET A. BODEN

Recognition: A Study in the Philosophy of Artificial Intelligence. By Kenneth M. Sayre.
University of Notre Dame, Notre Dame, Indiana, 1965. Pp. xxiii + 12. $6.95.

OVER the past decade a great deal of money and time, and a not inconsiderable amount of thought have been devoted to the problem of mechanically emulating one of the simpler skills of which we are capable. This is the ability to observe and describe the significant structure of pictures. For example, an electrical engineer might remark of a circuit diagram, 'the line connecting the grids of these two valves should not be present'. A physicist commenting upon an etched crystal surface viewed under a microscope might begin, 'there are lots of rectangles lying in pairs end to end . . .'. By contrast a child might say of an illustration in a reading book 'A'. The first two examples are concerned with the relationship between parts of the picture and are essentially *descriptive*. The verbal behaviour consists in a *string* of words, it is the grammatical and semantic relations between these words which is the description of the picture. The third example applies to the picture as a whole and is essentially *classificatory*: the verbal string exhibits no significant internal structure. It is largely with the problem of classification that Artificial Intelligence has been concerned: it is also the main burden of Professor Sayre's book. It will be clear from the first two examples, however, that letter recognition is not a representative sample of man's visual skills: a book which undertakes to examine the philosophy and practice of Artificial Intelligence and fails to bring out this fact is therefore disappointing.

Attempts to redress the balance [1] are rapidly leading to formal systems for the description of pictures which are closely akin to the 'generative grammars' formulated by the algebraic linguist as descriptive models of natural language. The outcome of this (if successful) will be computer programmes capable not only of classifying simple patterns but also of making descriptive remarks about pictures (cf. the electrical engineer) or of verifying that remarks about pictorial structure are true of some subset of a collection of pictures presented to the machine. The formulation of computer programmes of this sort (which Kirsch dubs 'Picture Language Machines') presupposes the existence of (1) an adequate method of picture analysis, such as will recover from a picture its content and structural organisation, (2) a corresponding language mechanism which would permit the generation and analysis of sentences having an appropriate content and form, and (3) a mechanism or mechanisms for mapping from one to the other. This tripartite division is typical of the way in which mechanisms of this complexity have to be organised for logical, heuristic or economic reasons. The simulation of behaviour is in part concerned to discover whether, for example, such divisions have counterparts in the organisation of human skills. In general this will involve a study of the performance both of the machine and of human subjects when carrying out similar tasks.

---

[1] R. A. Kirsch, 'Computer Interpretation of English Text and Picture Patterns', *I.E.E.E. Trans. on Electric Computers* 1964, EC-*13*, 4, 363-376

In Professor Sayre's book the attempt is made to reverse this line of inquiry by examining linguistic behaviour with a view to discerning the underlying perceptual or 'recognitive' organisation (Professor Sayre's term). The method can be represented as an analysis of subjective reports of recognitive behaviour. The sentence 'I am finding my wallet', is contrasted with 'I am looking for my wallet', 'I have found my wallet'. That the first is not a satisfactory English sentence provides evidence we are told for a distinction between 'process' verbs (scanning, watching, locking) on the one hand, and 'attainment' verbs (perceiving, discerning, discriminating) on the other. Moreover Professor Sayre argues that it provides information about the organisation of our visual skills by indicating that they rest upon 'processes' and 'attainments'. Subdivisions of these factors, e.g. 'perceptual' as against 'decisional' attainments may also be established by similar methods. The main criterion by which these sentences are rejected is that of 'common usage' with occasional appeal to grammaticality: in this way we are led to believe that a rigorous examination of a suitably chosen set of remarks will lead us to a correct model of recognitive behaviour. In fact, however, it can do no more than expose the set of concepts in terms of which we habitually describe our own behaviour. If 'attainment' verbs for example are *never* used with the present progressive, we may only conclude that there is wide agreement between language users on at least one aspect of the system they are describing with these sentences, i.e. the Picture Language Machine the speaker imagines himself to be. Whether this image is accurate can only be examined by experimental studies and digital computers are proving to be powerful tools in this connection. From this standpoint, 'analytic psychology' as Professor Sayre calls his method, is simply a vehicle for expressing our prior intuitions about our own organisation, it is not as he would have us believe (p. vii) 'an instrument of discovery for the simulation specialist'.

Independently of this analysis Professor Sayre subsequently proposes a model of letter recognition viewed as a communication task. He shows that classification is primarily concerned with the decoding operation, whereas recognition can be defined only by taking into account both source and receiver, that is by viewing recognition as 'a transaction between the penman and the perceiver of the symbols in point'. Of the decoding operation itself, Professor Sayre does not do the literature justice. He reviews only a limited category of approaches to the problem, and himself adds to the long list of untried, unproven, and essentially unworkable systems by proposing a new method of letter recognition based upon an inventory of 'basic' penstrokes.

As an account of the state of the art in Pattern Recognition this book is quite inadequate, it will not, I hope, deter other philosophers from taking an interest in this area of research.

MAX CLOWES

RECENT PUBLICATIONS ON THE PHILOSOPHY OF SCIENCE

(*a*) Books Received for Review

Alexander, E., *Matthew Arnold and John Stuart Mill*, Routledge & Kegan Paul, London, 1965, pp. 313, 45s.

Barfield, O., *Unancestral Voice*, Faber & Faber, London; 1965, pp. 163, 30s.

Barry, B., *Political Argument*, Routledge & Kegan Paul, London, The Humanities Press, New York, 1965, pp. xviii+364, 50s.

Bernal, J. D., *Science in History*, 3rd ed., C. A. Watts, London, 1965, pp. 1039, £4 4s.

Bolacchi, G., *Metodologia delle Scienze Sociali*, Edizioni Richerche, Roma, 1963, pp. xvi+241

Bonner, J. T., *The Ideas of Biology*, Methuen, London, 1965, pp. 240, 15s.

Bunge, M., *The Critical Approach to Science and Philosophy*, The Free Press of Glencoe, 1964, pp. xv+480

Caws, P., *The Philosophy of Science*, van Nostrand, Princeton, 1965, pp. xii+354.

Conn, G. K. T. and Turner, H. D., *The Evolution of the Nuclear Atom*, Iliffe Books, London; American Elsevier Publishing Co, Inc. New York, 1965, pp. 266, 55s.

Danto, A. C., *Analytical Philosophy of History*, Cambridge University Press, 1965, pp. xi+318, 55s.

Dobzhansky, T., *Heredity and the Nature of Man*, George Allen & Unwin, London, 1965, pp. x+179

Dooyeweerd, H., *Philosophy and Christianity*, North-Holland Publishing Co., Amsterdam, 1965, pp. 462

Eccles, J., *The Brain and the Unity of Conscious Experience*, Cambridge University Press, 1965, pp. 45, 4s. 6d.

Emmet, E. R., *Learning to Philosophise*, Longmans, Green & Co. London. 1964, pp. x+246, 13s. 6d.

Emmet, E. R., *Learning to Think*, Longmans, London, 1965, pp. x+168, 7s. 6d.

Emmet, E. R., *Companion to Learning to Think*, Longmans, London, 1965, pp. 82, 6s.

Erdey-Grúz, T. and Trencsényi-Waldapfel, I., *Science in Hungary*, Corvina Press, Váci Urca 12, Budapest, V. Hungary, pp. 316

Garnett, A. C., *The Perceptual Process*, University of Wisconsin Press, Madison, 1965, pp. 104, $3.75

Gauss, K. F., *General Investigations of Curved Surfaces*, Raven Press, New York, 1965, pp. 119, $3.95

Grossman, R., *The Structure of Mind*, University of Wisconsin Press, Madison, 1965, pp. viii+248, $6.00

Hacking, I., *Logic of Statistical Inference*, Cambridge University Press, 1965, pp. ix+232, 40s.

Hanson, N. R., *Patterns of Discovery*, Cambridge University Press, 1965, pp. 241, 12s. 6d.

Hempel, C. G., *Aspects of Scientific Explanation and other essays in the Philosophy of Science*, Free Press, New York; Collier-Macmillan, London, 1965, pp. x+505, 95s.

Hirst, R. J., *Perception and the External World*, Macmillan Co., New York; Collier-Macmillan, London, 1965, pp. viii+310, $1.95

RECENT PUBLICATIONS

Jaspers, K., *Leonardo, Descartes, Max Weber, Three Essays*, Routledge & Kegan Paul, London, pp. viii+274, 30s.

Jaspers, K., *Nietzsche*, University of Arizona Press, 1965, pp. 490, $12

Jones, R. F., *Ancients and Moderns*, University of California Press, Berkeley and Los Angeles, 1965, pp. xiii+345, 17s.

Kelly, E., *Ultrasonic Energy*, University of Illinois Press, Urbana, 1965, pp. 387, $12.50

Koyré, A., *Newtonian Studies*, Chapman & Hall, London, 1965, pp. viii+288, 50s.

Landé A., *New Foundations of Quantum Mechanics*, Cambridge University Press, 1965, pp. xii+171, 37s. 6d.

Larson, D. B., *New Light on Space and Time*, North Pacific Publishers, Portland, Oregon, 1965, pp. vi+264, $6.00

Lear, J., *Kepler's Dream*, University of California Press, 1965, pp. x+182, 40s.

Lemmon, E. J., *Beginning Logic*, Nelson, London, 1965, pp. x+225, 21s.

Letwin, S. R., *Pursuit of Certainty*, Cambridge University Press, 1965, pp. vii+391, 52s. 6d.

Levey, M. and Petruck, M., *Kushyar Ibn Labban: Principles of Hindu Reckoning*, University of Wisconsin Press, Madison, 1965, pp. xiv+114, $6.00

Mates, B., *Elementary Logic*, Oxford University Press, London, 1965, pp. x+227, 42s.

Merleau-Ponty, J., *Cosmologie du XX^{e} siècle*, Bibliothèque des Idées; éditions Gallimard, 1965, pp. 533

Moore, E. C., *William James*, Washington Square Press, New York, 1965, pp. x+194

Moritz, P. A., *Man the Universe and Mind*, Martigan Publications, Somerset, pp. xii+95.

North, J. D., *The Measure of the Universe.* A History of Modern Cosmology, Clarendon Press, Oxford, 1965, pp. xxviii+436, 70s.

Palmer, W. G., *A History of the Concept of Valency*, Cambridge University Press, 1965, pp. viii+178, 42s.

Pasquinelli, A., *Nuovi principi di epistemologia, Feltrinelli*, Milan, 1964, pp. 212

Ramsey, I., *Biology and Personality*, a symposium, Blackwell, Oxford, 1965, pp. vi+214, 30s.

Reichenbach, H., *Philosophic Foundations of Quantum Mechanics*, Cambridge University Press, 1965, pp. x+177, 12s. 6d.

Remits, E. L., *The Law in Nature and the Dynamics of Openness*, Runge Press, Ottawa, Canada, 1965, pp. iv+96

Royce, J., *Principles of Logic*, Book Sales Inc., New York City, 1961, pp. 77, $2.75

Shaftsbury, *Los Moralistas* (*The Moralists*, Trans, D. A. Sampietro), Instituto de Filosofia, Universidad Nacional de la Plata 1964, pp. 186

Simmons, E. D. *Essays on Knowledge and Methodology*, Ken Cook Co. Publishers, Milwaukee, Wisconsin, 1965, pp. v+145

Smith, V. E., *Science and Philosophy*, The Bruce Publishing Co. Milwaukee, 1965, pp. xiii+266, $5.50

Smythies, J. R., *Brain and Mind*, Routledge & Kegan Paul, London; The Humanities Press, New York, 1965, pp. x+272, 40s.

Spirin, A. S., *Macromolecular Structure of Ribonucleic Acids*, Reinhold, New York, 1964, pp. x+210, 80s.

Steinmetz, M. and Harig, G., *Bedeutende Gelehrte in Leipzig*, 2 vols, Karl-Marx-Universität, Leipzig, 1965, pp. xxxv+303; ix+211

Szabo, A. Gy., *Man and Law*, Akadémiai Kaido, Publishing House of the Hungarian Academy of Sciences, Budapest, 1965, pp. 206, 35s.

Tarski, A., *Introduction to Logic and the Methodology of the Deductive Sciences*, Oxford University Press, New York, 1965, pp. ix+252, 11s. 6d.

Thorpe, W. H., *Science, Man and Morals*, Methuen, London, 1965, pp. xii+176, 25s.

Topitsch, E., *Logik der Sozialwissenschaften*, Kiepenheuer & Witsch, Köln; Berlin, 1965, pp. 568

Vallenilla, E. M., *El Problema de la Nada en Kant*, Revista de Occidente, Madrid, 1965, pp. 304, 90 ptas

Waismann, F., *The Principles of Linguistic Philosophy*, Macmillan, London; St. Martin's Press, New York, 1965, pp. xii+422, 35s.

Weinberg, J. R., *Abstraction, Relation and Induction*, University of Wisconsin Press, Madison, 1965, pp. xii+156, $5.00

Whyte, L. L., *Internal Factors in Evolution*, Tavistock, Publications, London, 1965, pp. xxiii+81, 18s.

White, M., *Foundations of Historical Knowledge*, Harper & Row, New York and London, 1965, pp. ix+299, $6.95

Williams, L. P., *Michael Faraday. A Biography*, Chapman & Hall, London, 1965, pp. xvi+531, 70s.

Wittgenstein, L., *Philosophische Bemerkungen*, Blackwell, Oxford, 1965, pp. 348, 35s.

Woolf, H., *Science as a Cultural Force*, Johns Hopkins Press; Oxford University Press, London, 1965, pp. viii+110, 32s.

(*b*) Articles

General

Agassi, J., 'The confusion between physics and metaphysics in the standard histories of science', offprint from *Ithaca*-26, viii-2, ix, 1962, 231-250

Bonnett de Viller, A., 'Science et philosophie', *Dialectica*, 1965, **19**, 91-135

Buchdahl, G., 'A revolution in the historiography of science', *History of Science*, 1965, **4**, 55-69

Cheng, Chung-Ying and Resnik, Michael David, 'Ontic commitment and the empty universe', *J. Philos.* 1965, **62**, 359-364

Fackeldey, H., 'Essentia und esse und die seinsmomente des soseins und daseins', *Philos. Naturalis*, 1965, **9**, 114-137

Feibleman, J. K., 'Knowing, doing and being', *Ratio*, 1964, **6**, 12-27

Grünbaum, A., 'The bearing of philosophy on the history of science', *Science*, 1964, **143**, 1406-1412

Heintel, E., 'Galilei und Kant', *Philos. Naturalis*, 1965, **9**, 191-212

Hochberg, H., 'On being and being presented', *Philos. Sci.* 1965, **32**, 123-136

Körner, S., 'Philosophy, science and commonsense', *Ann. Jap. Ass. Philos. Sci.* 1961-65, **2**, 114-

Müller, H., 'Einheit als ziel des naturwissenschaft', *Philos. Naturalis*, 1964, **8**, 418-430

Robert, J. D., 'Science et méthodologie, Quelques réflexions', *Revue Philosophique de Louvain*, 1965, **63**, 116-125

Salmon, E. G., 'Mathematical roots of Cartesian metaphysics', *The New Scholasticism*, 1965, **39**, 158-169

Schorr, K. E., 'Logik und metaphysik (im bereich der Münsteraner schule-H. Scholz)', *Philos. Naturalis*, 1965, **9**, 46-54

Ströker, E., 'Philosophie als strenge wissenschaft', *Philos. Naturalis*, 1964, **8**, 377-396

Yukawa, H., 'Intuition and abstraction in scientific thinking', *Ann. Jap. Ass. Philos. Sci.* 1961-65, **2**, 94-

Epistemology

Achinstein, P., 'Models, analogies, and theories', *Philos. Sci.*, 1964, **31**, 328-350

Agassi, J., 'Discussion: analogies as generalizations', *Philos. Sci.*, 1964, **31**, 351-356

Anton, P., 'The status of the empirical principle', *Ann. Jap. Ass. Philos. Sci.*, 1961-65, **2**, 203-

Ayers, M. R., 'Counterfactuals and subjunctive conditionals', *Mind*, 1965, **74**, 347-364

Bunge, M., 'Physics and reality', *Dialectica*, 1965, **19**, 195-222

Feibleman, J. K., 'Material objects and the reference of signs', *Synthese*, 1963, **15**, 424-435

Gale, R. M., 'Why a cause cannot be later than it's effect', *Rev. Metaphysics*, 1965, **19**, 209-234

Gale, R. M. and Thalberg, I., 'The generality of predictions', *J. Philos.*, 1965, **62**, 195-210

Hacking, I., 'Salmon's vindication of induction', *J. Philos.*, 1965, **62**, 260-266

Hesse, M. B., 'Aristotle's logic of analogy', *Philos. Quart.*, 1965, **15**, 328-340

Holton, G., 'Presupposition in the construction of theories', from *Science as a Cultural Force*, ed. Harry Woolf, Johns Hopkins Press, 1964, 77-108

Krieger, L., 'Kant and the crisis in natural law', *J. Hist. Ideas*, 1965, **26**, 191-210

Laszlo, E., 'Is the primary-language an object-language?', *Foundations of Language*, 1965, **1**, 157-170

Leinfellner, W., 'Die konzeption der analytizität in wissenschaftlichen theorien', *Philos. Naturalis*, 1964, **8**, 397-417

Mellor, D. H., 'Experimental error and deducibility', *Philos. Sci.*, 1965, **32**, 105-122

Negel, T., 'Physicalism', *Philos. Rev.*, 1965, **74**, 339-356

Rack, K., 'Erkenntnistheorie als gemeinschaftsaufgabe von philosophie und naturwissenschaft', *Philos. Naturalis*, 1965, **9**, 153-162

Sellars, W., 'The identity approach to the mind-body problems', *Rev. Metaphysics*, 1965, **18**, 430-451

Stove, D., 'Hume, probability and induction', *Philos. Rev.*, 1965, **74**, 160-177

Titze, H., 'Die bewirkung im kausalen vorgang', *Philos. Naturalis*, 1965, **9**, 138-152

Tornebohm, H., 'Evidential support', *Acta Philos. Fennica*, 1965, 18, 267-276

Yamanouchi, T., 'Analysis and synthesis in science', *Ann. Jap. Ass. Philos. Sci.*, 1961-65, **2**, 200-

Logic

Büchler, L., 'On generalizing logical systems', *ΑΠΕΙΡΟΝ episteme*, 1965, **9**, pp. 44

Curry, H. B., 'Logic as a mathematical science', *Ann. Jap. Ass. Philos. Sci.*, 1961-65, **2**, 131-

Hintikka, J., 'On a combined system of inductive logic', *Acta Philos. Fennica*, 1965, **18**, 21-30

Putnam, H., 'Craig's theorem', *J. Philos.*, 1965, **62**, 251-260

Probability and Cybernetics

Bonsack, F., 'En quel sens les probabilités sont-elles subjectives?', *Dialectica*, 1965, **19**, 329-344

Caianello, E. R., 'Prospettive della cibernetca', *Scientia*, 1965, **100**, 118-128

Gonseth, F., 'Sur la Méthodologie du calcul des probabilités', *Dialectica*, 1965, **19**, 313-328

Heistermann, W., 'Modell und urbild-philosophische probleme der kybernetic', *Philos. Naturalis*, 1965, **9**, 22-45

Hesse, M., 'Analogy and confirmation theory', *Philos. Sci.*, 1964, **31**, 319-327

Salmon, J., 'What happens in the long run?', *Philos. Rev.*, 1965, **74**, 373-378

Salmon, W. C., 'The status of prior probabilities in statistical explanation', *Philos. Sci.*, 1965, **32**, 137-146

Sawada, N., 'Logic, cybernetics and philosophy', *Ann. Jap. Ass. Philos. Sci.*, 1961-65, **2**, 237-

Tillman, F. A., and Russel, B. R., 'Language, information and entropy', *Logique et Analyse (Louvain)*, 1965

Vickers, J. M., 'Some remarks on coherence and subjective probability', *Philos. Sci.*, 1965, **32**, 32-38

Will, F. L., 'The preferability of probable beliefs', *J. Philos.*, 1965, **62**, 57-67

Biological Sciences

Apolin, A., 'Die geschichte des ersten und zweiten hauptsatzes der wärmetheorie und ihre bedeutung für die biologie', *Philos. Naturalis*, 1964, **8**, 431-458

Cochran, A., 'Life and the wave properties of matter', *Dialectica*, 1965, **19**, 290-312

Lehman, H., 'Functional explanation in biology', *Philos. Sci.*, 1965, **32**, 1-20

Nooney, G. C., 'Mathematical models, reality and results', *J. Theoretical Biology*, 1965, **9**, 239-252

Raab, F. V., 'Of minds and molecules', *Philos. Sci.*, 1965, **32**, 57-72

Physical Sciences and Mathematics

Costa de Beauregard, O., 'Le "paradoxe" des corrélations d'Einstein et de Schrödinger et l'épaisseur temporelle de la transition quantique', *Dialectica*, 1965, **19**, 280-289

Destouches, J. L., 'General mathematical physics and schemas, application to the theory of particles', *Dialectica*, 1965, **19**, 345-348

Ellis, B., 'On the nature of dimensions', *Philos. Sci.*, 1964, **31**, 357-380

Emch, G. et Jauch, J. M., 'Structures logiques et mathématiques en physique quantique', *Dialectica*, 1965, **19**, 259-279

Fock, V., 'La physique quantique et les idéalisations classiques', *Dialectica*, 1965, **19**, 223-245

Graves, J. C. and Roper, J. E., 'Measuring measuring rods', *Philos. Sci.*, 1965, **32**, 39-56

Grünbaum, A., 'Carnap's views of the foundations of geometry', reprinted from *The Philosophy of Rudolf Carnap*, ed. P. A. Schilpp, 1963, The Library of Living Philosophers, Inc., 599-684

Hennemann, G., 'Bemerkungen zur entwicklung des physikalischen kraftbegriffes', *Philos. Naturalis*, 1964, **8**, 462-467

Hübner, K., ‘Zur gegenwürtigen philosophischen diskussion der quantenmechanik’, *Philos. Naturalis*, 1965, **9**, 3-21

Juhos, B., ‘Das prinzip der vitruellen geschwindigkeiten’, *Philos. Naturalis*, 1965, **9**, 55-113

Kunsemüller, H., ‘Zur axiomatik der quantenlogik’, *Philos. Naturalis*, 1964, **8**, 363-376

Kurth, R., ‘Uber das Newtonsche bewegungsgesetz’, *Philos. Naturalis*, 1964, **8**, 459-461

Landé, A., ‘Non-quantal foundations of quantum mechanics’, *Dialectica*, 1965, **19**, 349-357

Meurers, J., ‘Galilei und die naturwissenschaft der gegenwart’, *Philos. Naturalis*, 1964, **8**, 349-362

Park, D., ‘The idea of a particle in microphysics’, *Dialectica*, 1965, **19**, 246-258

Rositani, L., ‘Nuovi indirizzi nello studio della visione. E crollata la teoria della dualità?’, *Scientia*, 1965, **99**, 262-266

Schlegel, R., ‘The problem of infinite matter in steady-state cosmology’, *Philos. Sci.*, 1965, **32**, 21-31

Sciama, D. W., ‘The structure of the universe’, *Scientia*, 1965, **100**, 1-4

Stenius, E., ‘On the number of dimensions of physical space’, *Acta Philos. Fennica*, 1965, **18**, 227-241

Yanase, M., ‘Some remarks on concepts in theoretical physics’, *Ann. Jap. Ass. Philos. Sci.*, 1961-65, **2**, 280-

Behavioural and Social Sciences

Brand, B., ‘Critical reflections on behaviourism’, *Proc. Am. Philos. Soc.* 1964, **109**, 22-28

Büchler, L., ‘Logica etico-relativistica’, *ΑΠΕΙΡΟΝ episteme*, 1965, **5**, pp. 44

Büchler, L., ‘Psychological ground of the logic’, *ΑΠΕΙΡΟΝ episteme*, 1965, **11**, pp. 17

Granger, G., ‘Un problème d’axiomatisation en psychologie’, *Logique et Analyse (Louvain)*, 1965

Landesman, C., ‘The new dualism in the philosophy of mind’, *Rev. Metaphysics*, 1965, **19**, 329-345

Rorty, R., ‘Mind-body identity, privacy, and categories’, *Rev. Metaphysics*, 1965, **19**, 24-54

Thompson, J. W., ‘Prediction in physics and the social sciences’, *General Systems*, 1964, **9**, 15-23

Thompson, J. W., ‘Sociometry and the physical sciences’, part 1, *General Systems*, 1964, **9**, 1-5

Thompson, J. W., ‘Sociometry and the physical sciences’, part 2, the transfer of measuring techniques. *General Systems*, 1964, **9**, 7-13

ABSTRACTS

Philosophy of Science, 1964, 31 (4)

Norwood Russell Hanson, 'Stability Proofs and Consistency Proofs: a Loose Analogy'

A loose analogy relates the work of Laplace and Hilbert. These thinkers had roughly similar objectives. At a time when so much of our analytic effort goes to distinguishing mathematics and logic from physical theory, such an analogy can still be instructive, even though differences will always divide endeavours such as those of Laplace and Hilbert.

Brian Ellis, 'On the Nature of Dimensions'

In the first part of this paper it is shown that unit names, whether simple or complex, whether of fundamental, associative or derivative measurement, may always be regarded as the names of scales. In the second it is shown that dimension names, whether simple, like '$[M]$', '$[L]$' and '$[T]$', or complex dimensional formulae, may always be regarded as the names of classes of similar scales. Thus, a new foundation for the theory of dimensional analysis is provided, and in the light of this, its nature and scope are examined. Dimensional analysis is shown to depend upon certain conventions for expressing numerical laws.

Peter Achinstein, 'Models, Analogies, and Theories'

Recent accounts of scientific method suggest that a model, or analogy, for an axiomatised theory is another theory, or postulate set, with an identical calculus. The present paper examines five central theses underlying this position. In the light of examples from physical science it seems necessary to distinguish between models and analogies and to recognise the need for important revisions in the position under study, especially in claims involving an emphasis on logical structure and similarity in form between theory and analogy. While formal considerations are often relevant in the employment of an analogy they are neither as extensive as proponents of this viewpoint suggest, nor are they in most cases sufficient for allowing analogies to fulfil the roles imputed to them. Of major importance, and what these authors generally fail to consider, are *physical* similarities between analogue and theoretical object. Such similarities, which are characteristic in varying degrees of most analogies actually employed, play an important role in affording a better understanding of concepts in the theory and also in the development of the theoretical assumptions.

Mary Hesse, 'Analogy and Confirmation Theory'

The argument from analogy is examined from the point of view of Carnap's confirmation theory. It is argued that if inductive arguments are to be applicable to the real world, they must contain elementary analogical inferences. Carnap's system as originally developed (the λ-system) is not strong enough to take account of analogical arguments, but it is shown that the new system, which he has announced but not published in detail (the η-system), is capable of satisfying the conditions of inductive analogy. Finally it is shown that an elementary analysis of analogical inference yields postulates of the η-system with a minimum of arbitrary assumptions.

Philosophy of Science, 1965, 32(1)

Hugh Lehman, 'Functional Explanation in Biology'

This paper is concerned with the problem of giving a correct analysis of function statements as they are used in biology. Examples of such statements are (1) The function of the myelin sheath is to insulate the nerve fibre and (2) The function of chlorophyll is to enable photosynthesis to take place. After criticising analyses of such statements developed by Braithwaite, Nagel and Hempel an analysis is presented by the author. Finally the question of whether function statements are explanations is discussed briefly.

Richard Schlegel, 'The Problem of Infinite Matter in Steady-state Cosmology'

The creation-of-matter hypothesis of the Bondi-Gold-Hoyle steady-state cosmology requires that in an infinite time to which the first transfinite number may be assigned the number of atoms of matter produced would be equal to the cardinal number of the set of mathematical points in the continuum. The existence of a set of finite atoms with that cardinal number is physically unacceptable. The argument for the production of a non-denumerable set of atoms, in infinite time, is given in terms of a model which is shown to be isomorphic with the original Cantor 'diagonal' proof for the existence of a non-denumerable infinity. An alternative model which meets the requirements of the steady-state theory is presented; in this model, the number of atoms is explicitly no greater than countably infinite, and remains countably infinite as long as the past time of the universe is restricted to the unlimited set of finite unit-time intervals. If the origin of the steady-state universe is taken as being within that infinite set, expressed by the negative natural numbers, the contradiction of an atom at every mathematical point does not arise. The contradiction does arise if the origin is not within the set of finite numbers, and accordingly there is a restriction as to which concept of infinite past may properly be maintained in the steady-state theory.

John M. Vickers, 'Some Remarks on Coherence and Subjective Probability'

The interpretation of the calculus of probability as a logic of partial belief has at least two advantages: it takes the assignment of probabilities plausible in cases where classical frequentist interpretations must find such assignments meaningless, and it gives a clear meaning to partial belief and to consistency of partial belief.

John C. Graves and James E. Roper, 'Measuring Measuring Rods'

In this paper, we show that a restricted form of time travel both accords with special relativity kinematics and avoids several *prima facie* objections. We argue that such time travel provides a reasonable way to interpret certain phenomena which can readily be described, and the analogues of which have already been observed at the level of elementary particle reactions. We then describe how a time-travelling object could measure itself, and demonstrate how, in the appropriate circumstance, such an experiment could convince a theorist who insisted on a single criterion for length measurements that his standard had itself changed in length. The same experiment can shed light on the possibility of detecting a universal expansion; and we show that, given certain experimental results, one must conclude that every member of a class of time travellers has changed in length simultaneously, though perhaps only while going backwards in time.

Francis V. Raab, 'Of Minds and Molecules'

'Of Minds and Molecules' attempts to show the difficulties in mental-state brain-state monism. By exploring the differences in meaning between mental-state sentences and brain-state sentences, and by analysing the implications of the theory of the molecular composition of matter, a kind of dualism is arrived at that no scientist should feel uncomfortable with. It is a dualism without mental substance but it does not deprive mental states of their uniqueness. Arguments are given for the propriety of asserting causal connections and dependency relations between mental states and molecular states of the brain.

Philosophy of Science, 1965, 32(2)

D. H. Mellor, 'Experimental Error and Deducibility'

The view is advocated that to preserve a deductivist account of science against recent criticism, it is necessary to incorporate experimental error, or imprecision, in the deductive structure. The sources of imprecision in empirical variables are analysed, and the notion of conceptual imprecision introduced and illustrated. This is then used to clarify the notion of the acceptable range of a functional law. It is further shown that imprecision may be ascribed to parameters in laws and theories without rendering the deductive structure untestable. It is claimed that this analysis explicates the relation between laws and theories in a way that invalidates certain arguments against its deductive nature.

Herbert Hochberg, 'On Being and Being Presented'

Some philosophers have claimed that one must be *acquainted* with the elements of one's ontology. Also, believing that substrata and universals are required in an adequate ontology, these philosophers have claimed acquaintance with such objects. This paper attempts to analyse what is involved in such claims and to argue that they result from a number of confusions. The paper deals largely with the claim that substrata, or bare particulars, are presented since numerical difference is a simple fact that is presented. It further attempts to argue that while neither substrata nor universals, as such, are objects of acquaintance, there are some distinctions to be drawn between talk of universals as presented objects and talk of substrata being presented.

Wesley C. Salmon, 'The Status of Prior Probabilities in Statistical Explanation'

A consideration of some basic problems that arise in the attempt to provide an adequate characterisation of statistical explanation is taken to show that an understanding of the nature of scientific explanation requires us to deal with the philosophical problems connected with the nature of prior probabilities.

CAMBRIDGE

Concepts of Classical Thermodynamics

H. A. BUCHDAHL

Professor Buchdahl presents a systematic exposition of classical thermodynamics, against a background of general physical theory and on a purely phenomenological (i.e. non-statistical) level. The emphasis throughout is on meaning and physical significance. *Cambridge Monographs on Physics.* 45s *net*

Basic Concepts of Measurement

B. D. ELLIS

A general exposition of philosophical and logical principles of measurement. Mr. Ellis distinguishes and defines the basic concepts of measurement. He discusses the problem of classifying scales and special logical problems associated with each kind of scale. A translation of Mach's *Critique of the Concept of Temperature* is given as an appendix. 45s *net*

CAMBRIDGE UNIVERSITY PRESS

EVOLUTION AND MODIFICATION OF BEHAVIOUR

Konrad Lorenz

Dr. Lorenz helps to clear away the existing chaos concerning the conceptions of the 'innate' and 'learned' elements of behaviour. The technique of the deprivation experiment determining the source of information is extensively expounded. 25s

SCIENCE, MAN AND MORALS

W. H. Thorpe

'Prof. Thorpe assembles an impressive array of evidence, mathematical as well as biological, to prove the fantastic improbability of the idea that the origin of life, or even the simultaneous formation of two or more molecules of any given enzyme, occurred purely by chance.' *Nature.* 25s

METHUEN 11 New Fetter Lane, London, E.C.4

FOUNDATIONS OF LANGUAGE

International Journal of Language and Philosophy

Editorial Board Morris Halle (MIT), Peter Hartmann (Münster/W), K. Kunjunni Raja (Madras), Benson Mates (Univ. of California), J. F. Staal (Amsterdam), Pieter A. Verburg (Groningen), John W. M. Verhaar (Manila)

VOLUME II, No. 2

Articles B. Brainerd: Grammars for number names — Gilbert Harman: About what an adequate grammar could do — David E. Pfeifer: The question of reference in the writings of J. A. Fodor and J. J. Katz — Paul M. Postal: Review article (A. Martinet. *Elements of general linguistics*)

Reviews P. A. Verburg. Taal en functionaliteit (W. Z. Shetter)

Anna-Teresa Tymieniecka (ed.): *Contributions to logic and methodology in honor of J. M. Bocheński* (William Kneale)

Robert E. Longacre. *Grammar discovery procedures. A field manual* (Peter A. M. Seuren)

Yehoshua Bar-Hillel: *Language and information. Selected essays on their theory and application* (Maurice Gross)

Benjamin Elson and Velma Pickett: *An introduction to morphology and syntax* (Peter A. M. Seuren)

Publications Received for Review.

FOUNDATIONS OF LANGUAGE is published quarterly
Annual subscription price ƒ 42.—($11.75; 84s.)

D. REIDEL PUBLISHING COMPANY / DORDRECHT - HOLLAND

JOURNAL OF THE HISTORY OF IDEAS

A Quarterly devoted to Intellectual History

VOLUME XXVII JULY—SEPT. 1966 NUMBER 3

TABLE OF CONTENTS

ARTICLES

The Classical Greek Concept of Slavery . . . . *Victoria Cuffel*
Majority History and Post-Biblical Jews . . . . . *Gavin I. Langmuir*
Trust and Deceit in Machiavelli's Comedies . . . . . . *Martin Fleisher*
Perennial Philosophy: from Agostino Steuco to Leibniz . . *Charles B. Schmitt*
Anarchism and anti-Intellectualism in Russia . . . . . . *Paul Avrich*

NOTES

From Constance to 1688 Revisited . . . . *Francis Oakley*
The Trinitarian Framework of John Donne's Holy Sonnet XIV . *Lucio P. Ruotolo*
The Pilot of the Galilean Lake . . . . *Ernest Tuveson*
An Analogue of Transcendentalism . . . . *John B. Wilson*

Subscription Rates (outside U.S.A. post paid): $6.50 for one year,
$1.75 for single issue

Payable to

JOURNAL OF THE HISTORY OF IDEAS, INC.
The City College, New York, N.Y. 10031

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

JUNE 1966, VOL. XXVI, No. 4

CONTENTS

I. ARTICLES

Idealistic Social Philosophy and Bernard Bosanquet, by JOHN HERMAN RANDALL, Jr.
The Implications of Experience for Psychological Theory, by W. W. MEISSNER, S.J.
On Characterizing the Psychological, by HERBERT HEIDELBERGER
Tastes, by JOHN O. NELSON
Social Perception and Moral Judgment, by GUSTAV ICHHEISER
Learning the Language, by JOHN M HEMS

II. DISCUSSIONS

Behaviorism and Phenomenology, by V. J. McGILL

III. REVIEWS, NOTES AND NEWS, AND RECENT PUBLICATIONS

Subscription prices $6 00 per year for individuals; $7.50 per year for institutions. Separate issues, $1.65 and $2.00 respectively. Postage added for back issues only.

A circular listing the contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo, Buffalo, New York, 14214

THE REVIEW OF METAPHYSICS

A PHILOSOPHICAL QUARTERLY

Edited by RICHARD J. BERNSTEIN

Volume XX, No. 1 *Issue* 77 *September* 1966

ARTICLES·	
Robert R. Ehman	*Subjectivity and Solipsism*
Ivor Leclerc	*Kant's Second Antinomy and the Metaphysics of Whitehead*
Robert L. Holmes	*John Dewey's Moral Philosophy in Contemporary Perspective*
CRITICAL STUDIES:	
W. K. Wimsatt	*Sparshott on Aesthetics: A Guided Tour*
Carl R. Hausman	*Understanding and the Act of Creation*
Richard J. Bernstein	*Sellars' Vision of Man-in-the-Universe, I*
BOOKS RECEIVED:	
Edward A. Reno, Jr. and Staff	*Summaries and Comments*

PHILOSOPHICAL ABSTRACTS
DOCTORAL DISSERTATIONS, 1966
VISITING PHILOSOPHERS FROM ABROAD
ANNOUNCEMENTS

As of September 1, 1966 our address is
HAVERFORD COLLEGE, HAVERFORD, PENNSYLVANIA 19041 USA

Annual Subscription Rates: Individuals $6.00; Students $4.00

Readers resident in Great Britain may subscribe at a special rate through our British representative, Professor Stephan Körner, Department of Philosophy, The University, Bristol 8, England

SYNTHESE

An International Journal for Epistemology, Methodology and Philosophy of Science

Editor: JAAKKO HINTIKKA
University of Helsinki and Stanford University

Board of Consulting Editors: David M. Armstrong, Yehoshua Bar-Hillel, Arthur W. Burks, Donald Davidson, Hans Freudenthal, Béla Juhos, Jerrold J. Katz, Imre Lakatos, G. Nuchelmans, Wesley C. Salmon, Erik Stenius, Patrick Suppes, Klemens Szaniawski, H. Törnebohm.

Contents of Volume XVI, No. 1 September, 1966

Proceedings of the E. W. Beth Memorial Colloquium, Paris, 19-21 May, 1964 (with additions)
Memorial Address by *G. H. von Wright.*
Allocution by *J.-L. Destouches.*
Beth's Tableau Method *J. J. F. Nieland.*
Quelques remarques sur les "Tableaux de Beth" by *Marcel Guillaume.*
Une hypothèse sur l'extension des relations finies et sa vérification dans certaines classes particulières (deuxième partie) by *Roland Fraïssé.*
Remarques sur la théorie intuitioniste des espaces linéaires by *A. Heyting.*
La philosophie géometrique de Henri Poincaré by *J. J. A. Mooij.*
Logique et théories physiques by *J.-L. Destouches.*
L'argument probabilistique pour une logique non-classique de la mécanique quantique by *Patrick Suppes.*
Conclusions by *A. Heyting.*
Bibliography of E. W. Beth by *J. F. Staal.*
Reviews
Béla Juhos, Elemente der Physikalischen Semantik by *Herbert Schleichert.*
Howard E. Smokler, Aspects of Scientific Explanation and Other Essays in the Philosophy of Science by *Carl G. Hempel.*
Books Received

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND

Annual Subscription Price Dfl. 48—post-free

The Journal of Philosophy

A BIWEEKLY journal containing philosophical articles of current interest and encouraging the interchange of ideas.

Editors: John H. Randall, Jr., Sidney Morgenbesser, Arthur Danto, Charles D. Parsons, and James J. Walsh. $7.50 a year; to students, $6.00; 50 cents a copy.

NOW AVAILABLE: Cumulative *Fifty-year Index* 1904-1953; articles classified by subject and author; 452 p., cloth $12.00. Also, *Ten-year Supplement*, 1954-1963; $3.00 cloth, $2.00 paperbound. Both *Indexes*, bound together, $15.00.

BACK NUMBERS and volumes: 1960 to 1965 (LVII to LXII), 50¢/issue, $12/volume; 1940 to 1959 (XXXLVII to LVI), 75¢/issue, $18/volume; 1904 to 1939 (I to XXXLVI), $1/issue, $24/volume.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NEW YORK 10027

VOLUME VIII JUNE 1966 No. 1

RATIO

edited in co-operation with
H. B. Acton (London), P. Bernays (Zürich), A. Church (Princeton), H. Cherniss (Princeton), G. Henry-Hermann (Bremen), W. C. Kneale (Oxford), E. Nagel (Columbia), K. R. Popper (London) and J. W. N. Watkins (London)
by
Stephan Körner (Bristol)

W. STEGMÜLLER	Explanation, prediction, scientific systematization and non-explanatory information
W. BEISSWANGER	Hermann Weyl and Mathematical texts
N. RESCHER and J. ROBINSON	Temporarily Conditioned Descriptions
J. O. WISDOM	Testing a Psycho-Analytic Interpretation
H. KHATCHADOURIAN	About Imaginary Objects
J. KNOX	On Mr. Nelson's Rejection of Sense-Data
A. FLEW	Did Hume Distinguish Pure from Applied Geometry?
Reviews:	
R. HARRÉ	*Matter and Method* (G. Buchdahl)
M. WARNOCK	*The Philosophy of Sartre* (D. Pole)

Price 15s. net **Annual Subscription: 27s. 6d. net**

BASIL BLACKWELL · OXFORD

ZEITSCHRIFT FÜR PHILOSOPHISCHE FORSCHUNG

in Verbindung mit der
"Allgemeinen Gesellschaft für Philosophie in Deutschland"
Herausgegeben von G. Schischkoff, München
Single copy $4; volume (four issues) $13.50

PHILOSOPHISCHER LITERATURANZEIGER

Ein Referateorgan für die Neuerscheinungen der Philosophie und ihrer gesamten Grenzgebiete, in Verbindung mit der "Zeitschrift für philosophische Forschung"
Herausgegeben von G. Schischkoff, München
Single copy $2.30; volume (six issues) $10.80

PHILOSOPHIA NATURALIS

Archiv für Naturphilosophie und die philosophischen Grenzgebiete der exakten Wissenschaften und Wissenschaftsgeschichte
Begründet von Eduard May
Herausgegeben von Joseph Meurers
Single copy $2,75; volume (four issues) $10

VERLAG ANTON HAIN KG., Meisenheim/Glan. (Germany)

CALCUTTA
CENTRAL LIBRARY.

Printed in Great Britain at The University Press, Aberdeen

The British Journal for the Philosophy of Science

BJPS

CALCUTTA UNIVERSITY
CENTRAL LIBRARY.

Cambridge University Press

November 1966

Volume 17 No. 3 15s. net

Full particulars of the

JOURNALS

published by the

CAMBRIDGE UNIVERSITY PRESS

may be had from

The Manager
Cambridge University Press
Bentley House 200 Euston Road
London, N.W.1

Requests for advertisement rates and data should be made to "Publication Dept." at the above address

---

ZEITSCHRIFT FÜR PHILOSOPHISCHE FORSCHUNG

in Verbindung mit der
"Allgemeinen Gesellschaft für Philosophie in Deutschland"
Herausgegeben von G. Schischkoff, München
Single copy $4; volume (four issues) $13.50

PHILOSOPHISCHER LITERATURANZEIGER

Ein Referateorgan für die Neuerscheinungen der Philosophie und ihrer gesamten Grenzgebiete, in Verbindung mit der "Zeitschrift für philosophische Forschung"
Herausgegeben von G. Schischkoff, München
Single copy $2.30; volume (six issues) $10.80

PHILOSOPHIA NATURALIS

Archiv für Naturphilosophie und die philosophischen Grenzgebiete der exakten Wissenschaften und Wissenschaftsgeschichte
Begründet von Eduard May
Herausgegeben von Joseph Meurers
Single copy $2,75; volume (four issues) $10

VERLAG ANTON HAIN KG., Meisenheim/Glan. (Germany)

Brit. J. Phil. Sci. (1966), Volume 17, Number 3, 177–204
Printed in Great Britain

THE *RAISON D'ÊTRE* OF INDUCTIVE ARGUMENT[*]

MAX BLACK

1 *Some definitions and assumptions*

BY an inductive argument I mean a set of declarative propositions (the premises), together with another declarative proposition (the conclusion), that purports to be supported by those premises *without following from them.* By definition, therefore, no inductive argument is a deductive argument. All of the inductive arguments I shall consider are intended to be fully explicit: the intended degree of support for the conclusion is supposed to be provided wholly by the stated premises, without assistance from unstated assumptions.[1]

A correct deductive argument is usually called 'valid'; to avoid confusion, a correct inductive argument may be called 'sound'. Since an inductive argument does not aspire—or should not aspire—to satisfy the standards of deductive correctness, it would be misleading to call it invalid. I shall be assuming, without argument, that some inductive arguments are sound and some are not.[2]

The foregoing definition deviates from the main traditions of the

---

[*] Received 3.i.66

[1] In practice, it is not always easy to distinguish an inductive argument, as here defined, from a deductive enthymeme. Consider the following example. *Pippa is a bird so Pippa has wings*: is this to count as a deductive argument or as an incomplete inductive one? If the suppressed premise is *All birds have wings*, the argument will have to count as a deductive enthymeme; if it should be *Most birds have wings*, it will be an inductive argument (of the sort called a 'proportional syllogism') or an invalid deductive one, depending upon whether the conclusion is or is not intended to follow demonstratively from the premises. If the argument is actually presented by a speaker available for interrogation, he can sometimes be asked how he intends the argument to be taken. I shall not consider any further difficulties involved in identifying inductive arguments.

[2] Anybody who denies this must presumably hold that it is no better to argue from *Pippa is a bird and Most birds have wings* to *Pippa has wings* than it would be to argue from the same premises to the contrary conclusion. This seems plainly absurd. The premises, by themselves, provide some reason for thinking that Pippa has wings and hence, some reason for not thinking that Pippa does not have wings. Even if the argument about Pippa be regarded as very weak, it should be recognised as sound, so far as it goes.

philosophy of induction in two ways. Classical writers take induction to be 'generalisation from experience' (Mill) or 'the logical processes employed in the inductive sciences, so far as these infer from particular facts the principles that explain them' (W. E. Johnson, *Logic*, **3**, 395), or something similar. However, I find it useful to consider all forms of 'non-demonstrative' argument together. Besides generalisation from particulars, inductive arguments, thus comprehensively viewed, include 'eduction' (Johnson), or 'argument from particulars to particulars', and 'direct inductive arguments' from imperfect generalisations to their instances.

In the second place, there is a powerful tradition in the philosophy of induction that is committed to viewing allegedly inductive arguments as special kinds of inexplicit deductions.[1] Mill, and many other writers influenced by him, have insisted that every argument that would commonly be called 'inductive' needs supplementation by some grand major premise concerning the uniformity of nature. Other writers claim that an inductive argument can be justified only if its conclusion is sufficiently weakened, by including explicit reference to probability, for the conclusion to be entailed by the given premises. If either demand is accepted, what is commonly called an inductive argument becomes a special case of a deductive one. By contrast, I wish to explore the consequences of treating induction, in the broad sense, as a distinctive mode of argument having its own special *rationale*, and subject to standards of correctness that are not those of deduction.

Accordingly, I set out the general form of a complete inductive argument as follows:

$$\begin{array}{c} P_1 \\ P_2 \\ \cdot \\ \cdot \\ \cdot \\ P_n \\ \textit{Therefore (probably):} \\ K \end{array}$$

Here, the qualifier 'probably' may be conceived to be attached, as shown to the 'Therefore' (the sign of illation); when the conclusion appears detached from its premises, it is useful to attach the ' probably ' qualifier to it (' Probably, K ') as a reminder.[1]

---

[1] 'What is called induction appears to me to be either disguised deduction, or a mere method of making plausible guesses' (Bertrand Russell, *Principles of Mathematics*, p. 11 n.).

The word ' probably ', which might be called a *reliability indicator*, has something of the force of ' more likely than not '.[2] Its occurrence shows that the argument is intended to be inductive, not deductive. Other reliability indicators in common use are ' very probably ', ' almost certainly ', ' the chances are ', and ' beyond all reasonable doubt '. In sophisticated contexts, such as those of statistical argument, the reliability in question may be indicated by some number or range of numbers.

It will be noticed that I am taking the conclusion of an inductive argument to be *detached* (although I am aware of the difficulties that may arise from this decision).[3] I wish to take the conclusion as K, *simpliciter*, without the intrusion of any indication of probability in the conclusion itself.[4] I shall not be able to discuss in this paper arguments whose conclusions are propositions that include reference *to* probabilities.

I propose now to consider the drawing of inductive *inferences*, i.e.

---

[1] In an ampler context, this might be a convenient place at which to develop a notion of *guarded assertion*.

[2] If the truth of the premises tips the balance in favour of the conclusion as against its negation only very slightly, one might hesitate to *assert* the favoured proposition. So, a more plausible expansion of ' Therefore (probably) ': might be ' Therefore (more likely than not *by a sufficient margin*) '.

[3] For discussion of these, see especially C. G. Hempel, ' Inductive Inconsistencies ', *Synthese*, 1960, 12, 439-469. One of Hempel's objections to ' detaching ' the conclusion is that contrary conclusions may need to be inferred from alternative sets of true premises (op. cit. p. 442). This seems to me a peculiarity of inductive inference that simply has to be accepted. More troublesome, I think, is the fact that K may be soundly inferred from P, and L from K, without L being a sound inference from P, i.e. the difficulty that the relation of inductive implication is apparently not transitive. Again P may sufficiently support K_1, and also sufficiently support K_2 without sufficiently supporting the conjunction $K_1 . K_2$. The notion of ' guarded assertion ', mentioned in footnote 1, above, offers some prospect of handling such difficulties.

[4] I would argue for the necessity for this by urging that only in this way can inductive inference have any bearing upon subsequent verification. If the conclusion is of the form *The probability of K (relative to P) is such-and-such*, its truth-value is independent of whatever proves to be the case (whether on the ' logical ' or on the ' frequency ' interpretations of probability). Attempts, such as those of Carnap, to provide a link between probability and verification—or, at least, between probability estimates and rational action—by relying on some principle for the maximising of utility are, in my opinion, unsuccessful. For the most that can be assured by any such principle is ' expected ' gain, i.e. the *probability* of some desired, theoretical or practical, outcome.

the affirmation of inductive conclusions because they are, or are taken to be, sufficiently supported by true premises of sound inductive arguments. I wish to view such inferences as falling within the scope of a *purposive, rule-governed, practice.*

To be more explicit: Let us call an episode, in which some inductive conclusion is actually drawn by somebody, an *inductive performance.* I conceive of each such performance as determining a class of arguments, all of whose members are appraised as sound or unsound in virtue of rules that determine what is to count as inductive correctness for the class of cases in question.[1] Various rules of this sort, in their totality and interaction, constitute a formalised *practice*—or, if the reader prefers, an institution, a tradition—that defines inductive correctness and prescribes associated modes of conduct. Such conduct I take to be, notably, a readiness to ' stand by ' an asserted inductive conclusion, to defend it against attack, to use it as a premise for further arguments, to accept it as a premise when it is offered by another person, and so on. A short formula for this complex maze of related activities might be the *public expression and elaboration of the consequences of belief.* The practice or institution of inductive inference is a social device for regulating and appraising the publication of beliefs purporting to be derived from other beliefs but not entailed by them.

2 *The problem stated*

Given this conception of inductive inference, I now feel obliged to seek some account of the *rationale* of the inductive practice and of the specific activities falling within its ambit. Induction, like the Sabbath, was made for man—and not vice versa. Accordingly, it should be possible to exhibit the overall purposes served, or intended to be served, by the encompassing practice, and the ways in which the rules and procedures enjoined by that practice contribute to the fulfilment of these purposes.

---

[1] Some will wish to object that inference has to do with belief—which is not a matter of choice or decision and, therefore, not subject to the control of rules. Judith Jarvis Thomson says, of a proposed rule of inference: 'To put it in the form of an imperative . . . would unfortunately be mad. Must I really be inferring statements from each other all the time? So also for attempts to set it out as a conditional imperative. Infer this from that, whenever—whenever I want to infer this from something? ' (' Reasons and Reasoning ', in Max Black, ed., *Philosophy in America,* 1965, p. 287). She complains that it is not clear what ' actions ' could be enjoined by rules of inference, since they must tell the reasoner ' to think something, and not merely to say or write something ' (op. cit. p. 289). The implied separation between thought and action seems hard to maintain. In any case, Mrs Thomson takes too narrow a view of the character of the rules in question. They could, reasonably, be taken as rules of evaluation, specifying when inductive arguments are to *count as* being correct. Whether or not given reasoners find themselves able to obey the rules is therefore irrelevant. So long as they can understand that certain patterns of thought have to be criticised as incorrect, that will suffice for the practical utility of such rules.

The peculiar urgency of this programme for exhibiting the *rationale* of induction can be shown by contrasting the case of deductive inference. There, the *rationale* seems straightforward enough, at first sight. When the premises of a deductive inference are true and the argument is valid (is correctly conducted according to the prescriptions of the deductive practice), the conclusion is also true: valid deduction is truth-preserving, truth-transmitting. Thus, from knowledge of the truth of premises, we can sometimes acquire explicit knowledge of the truth of other propositions implicated by those premises. The benefits are obvious. If we desire to increase available knowledge, without recourse to experience or some other external source, deduction will sometimes satisfy the desire. Deduction is like a dredge that brings hidden cognitive treasure to light—or like the cashing of a draft on a fund of potential knowledge. Given an antecedent interest in increasing our store of available knowledge, the point of deduction is clear enough and no further rationalisation is needed. (This sketch is, of course, incomplete, if only because we often wish to argue from the falsity of a conclusion to the falsity of one of its premises, or to draw the deductive consequences of premises only hypothetically entertained.)

Now contrast the case of inductive inference. Any inductive performance must be connected in some way with an antecedent interest in discovering the truth. But it is a striking feature of inductive inference, in which it differs importantly from deductive inference, that a sound inductive argument with true premises may have a false conclusion. Anybody who reasons deductively from true premises to a false conclusion must have made a mistake; but if he derives a false inductive conclusion from true premises, he may have been only unlucky. This makes the aim of inductive performances puzzling. A sceptic might reasonably ask: If the object of induction is to arrive at the truth, and if even the best inductive inference may fail to achieve this object, what *is* the point of such a performance or of the practice to which it belongs?

CALCUTTA CENTRAL LIBRARY.

3 *Some distinctions applicable to purposive activities*

A few distinctions, applicable to all purposive activities, may help us to find our bearings. I shall distinguish between a *performance*, the *activity* of which it is an instantiation, and the general *practice* encompassing the activity and determining its proper mode of execution. (Thus, the particular episode consisting of somebody's opening a game of chess by moving his King's pawn on a particular occasion will be his *performance*; the corresponding ' activity ' will be identified as *moving P-K4 on the first move*: and the enveloping ' practice ' is, of course, that of playing chess. An activity, in the sense here understood, is more abstract than a performance: the latter must occur at a specified place and time, while the former is, roughly speaking, the same as *what* is performed, or as I shall often say, ' instantiated '.)

I wish to distinguish, also, between the *immediate purpose* of the performance, the *general purpose* of the relevant activity, and the *global purpose* of the encompassing practice. (The immediate purpose of a dressmaker may be to put a stitch of a certain form in a particular place on the material; the general purpose of stitching may be that of holding pieces of material together in the desired shape; and the global purpose of dressmaking is, say, that of making garments that are comfortable and pleasing.) The threefold distinction I have introduced is somewhat rough-and-ready, but it will serve our purpose.

In the cases to be discussed, we may assume that the immediate purpose of a performance is *conducive* to the general purpose, and that the general purpose of an activity is *conformable* to the global purpose of the practice.

We may, with advantage, also introduce a cross-classification of purposes as *substantive* and *procedural*. Typically, the immediate purpose of a performance is to achieve a certain end-state or result (the substantive purpose) but also to do so in a certain fashion, mode, or style (the procedural purpose). (Sometimes, the ' procedural purpose ' alone is relevant—as when a pianist plays a sonata. Here there is no ' end state ' to be produced—no artifact to be generated.)

The procedural purpose is usually intended to be conducive to the substantive purpose. Thus, a golfer addresses himself to the ball in a certain way, with the hope that making the stroke in ' proper form ' will serve his immediate purpose of getting the ball as close to the hole as possible. When achievement of the subordinate procedural purpose is certain to lead to achievement of the immediate purpose, I shall say that the former is *strongly* or categorically conducive to the latter. (Thus the tactic of keeping contact with one bounding wall of a maze is strongly conducive to the immediate purpose of getting out of that maze, since it constitutes an achievement procedure for that task.) When achievement of the procedural purpose contributes to, but does not guarantee, achievement of the immediate purpose, I shall say that it is *weakly* conducive to it. (Thus, aiming a gun at the target is weakly, but only weakly, conducive to the purpose of hitting that target.)

To sum up: I am thinking of the particular inductive performance of drawing a conclusion from given premises as having an immediate purpose that is both substantive and procedural; I think of the corresponding inductive activity that is thereby instantiated as having some general cognitive purpose; and I assume that the immediate purpose is at least weakly conducive to the general purpose, while the latter is such as to contribute in some way, still to be determined, to the overall global purpose, whatever that may be, of the inductive practice.

4 *The problem restated in application to an example*

Consider, now, somebody who argues as follows:

Most businessmen are Conservatives
Grumbacker is a businessman
Therefore (probably):
Grumbacker is a Conservative.

I take the *activity* of which this performance is an instantiation to be that of arguing from *Most B are C and A is a B* to a conclusion of the form *A is a C*. I want to ask, first, whether the *performance* has an immediate substantive purpose and, if so, what that purpose is.

It seems to me quite unplausible to say, as some philosophers of induction do, that there is no immediate purpose, apart from the procedural one of performing correctly, i.e. of reasoning soundly. Those who take this position sometimes proceed to say that the only rational purpose of a particular inductive performance is that of providing a conclusion conducive to the general long-term purpose of the activity—or, perhaps, still more indirectly, to the global purpose of the encompassing practice. They are implying, I think, that so far as a particular inductive inference is concerned, all that the reasoner is immediately trying to achieve is a sound argument, with possible implications for wider purposes (behaving rationally on the whole—or sustaining a rational practice). This is no more plausible, I think, than contending that a man who gives a promise has no immediate interest in the outcome of his performance, but is solely interested in being judged to be a morally trustworthy person—or, is perhaps solely interested in maintaining the institution of promising. It seems to me clear enough that anybody, not a lunatic, who does perform an inductive inference is trying to achieve something immediately—and something more than reasoning soundly.

If there is an immediate substantive purpose, it could hardly be anything else but one of the following: either, to reach a conclusion that is true—or, alternatively, to reach one that is at least more likely than not to be true. Should the less ambitious of these alternatives be adopted as correct, the global purpose of the entire practice could not be anything more grandiose than that of arriving at a body of empirical belief that is on the whole well confirmed—is more likely than not to be true. For if nothing better than probable truth is to be expected of the individual conclusions, only probability or likelihood can be expected of their totality. Many able philosophers of science have acquiesced, or have positively gloried in, this deflationary conception of empirical inquiry, and they might conceivably be right. I shall, however, adopt the more interesting hypothesis that the immediate substantive purpose of an inductive performance is to arrive at a true conclusion. I believe, indeed, that anybody inductively reasoning from true premises has two immediate ends in view: to reason correctly and also to get a true conclusion. To be sure, all that even the best reasoner can guarantee is that he has made no mistake in reasoning, for whether his conclusion turns out to be true, as he hopes, is something that depends on more than his skill. (There is an element of risk in all inductive reasoning.) Nevertheless, to arrive by sound reasoning at a false conclusion must be judged to be a setback: the reasoner has failed in his immediate purpose, though through no fault of his own.

That the substantive purpose of getting a true conclusion may not be achieved is no objection to this view. When I play the patience game called Spider, I know that the chances of success are small, yet my object in playing each hand is correctly described as that of ' getting the cards out ' *this* time. So long as I know that success is not impossible, there is no absurdity in setting myself to perform a task that

I may fail to accomplish. So also in the case of the inductive inference. So long as the reasoner knows, as he does, that it is not impossible for him to reach a true conclusion of the sort he desires, it is quite sensible for him to aim at getting such a conclusion each time he tries. (The higher the likelihood of truth, the more confidently he may set himself in pursuit of it.)

The inductive reasoner tries to obtain a true conclusion, not by accident but by means of a sound argument. His procedural purpose is, therefore, simply that of reasoning soundly in accordance with the relevant rules governing such reasoning. From what has already been said, we can see that this procedural purpose could at best be *weakly* conducive to the substantive aim of getting a true conclusion, for the conclusion's truth is never guaranteed. But it is not obvious how it can be even weakly conducive. How does conformity to rules of sound inductive reasoning contribute to the immediate purpose of arriving at a true conclusion? In order to see this, we shall need to reflect upon the relation between the premises and the conclusion of a sound inductive argument.

5 *Relations between premises and conclusion of a sound induction*

If we unite the two premises of our specimen inductive argument, we get the amalgamated premise: *Most businessmen are Conservatives and Grumbacker is a businessman.* Call it *P*. What is the connection between this proposition and the conclusion, *Grumbacker is a Conservative* (call it *K*) that entitles us to say that the passage from the former to the latter is sound? I can think of the following two answers, and no others: (*a*) Belief in the truth of *P* is a good reason for belief in the truth of *K*; and (*b*) if *P* is true, then *K* is more likely than not to be true.

There is a sense in which both answers amount to very nearly the same. First, it seems to me that acceptance of (*b*) necessarily requires acceptance of (*a*): If you think that *K* is more likely than not to be true, given that *P* is true, you are obliged to hold that belief in the truth of *P* is, so far, a good reason for belief in the truth of *K*. In the second place, the converse also seems true: If you hold that belief in the truth of *P*, by itself, is a good reason for holding *K* to be true, you are obliged to hold that if *P* is true, *K* is more likely to be true than not. Moreover, the two conditionals seem to be necessary truths, depending, roughly speaking, upon what we mean by 'belief', 'good reason', and 'likelihood'. It is not a contingent matter of fact that 'good reason' and 'likelihood' are connected in the manner I have indicated; it is part of what we mean by the two expressions that they should be thus connected. (We could not make sense of somebody who urged that because *A* was less likely than not to be true, given *B*, *therefore* belief in the truth of *B*, by itself, would be a good reason for belief in the truth of *A*.) One difference between the two formulations is that talk about 'good reasons' belongs to a context of evaluation, with fairly direct consequences for practice, while talk about 'likelihood' is more theoretical, and does not trail practical consequences so obviously. To say something is a 'good reason' is to prod your hearer, while to speak merely of likelihood is to adopt a cooler, more dispassionate posture. (Cf. the difference between asking somebody to close a window and merely remarking that it would be nice to have the window closed.)

It follows, if I am not mistaken, that a question about the weak conductiveness of the procedural purpose of *this* kind of inductive performance can be answered with remarkably little trouble.

If *K* is the conclusion of a sound inductive argument from *P*, it is a necessary truth that *K* is more likely than not to be true, given *P*—and a necessary truth that belief in the truth of *P* is, by itself and so far, a good reason for belief in the truth of *K*. Hence, if follows at once that performing the inductive inference in correct form is weakly conducive to the desired substantive goal of getting a true conclusion. For that accomplishing *A* is *likely* to lead to the accomplishment of *B* is precisely what we mean by saying that *A* is weakly conducive to *B*. If reasoning correctly is *likely* to lead to a true conclusion, we have all that can reasonably be desired in seeking a link between the mode of performance and its desired end product.

This line of thought, although perfectly correct, will probably seem too glib to be satisfactory: it may well have the unsatisfying appearance of producing an invisible rabbit from a non-existent hat.

Consider the following interchange between a critic of this standpoint and its defender:

A. Given that I am primarily interested in getting a *K* that is true, why should I bother about arguing from *P* 'soundly'?

B. Because *P* is a good reason for *K*, so the truth of *P* makes it more likely than not that *K* is true. Thus *P* is conducive to the truth of *K*, which is what you want.

A. But how do we know that *P* is a good reason for *K*?

B. That is a necessary truth. To use 'good reason' correctly is, *inter alia*, to consent to call *this* kind of argument one that supplies a good reason for the conclusion. To do anything else would be to violate a linguistic practice to which you are as firmly committed as I am.

What B says is perfectly right, but it will be singularly unsatisfying to anybody who has a genuine philosophical perplexity about the status and rationale of inductive procedures. One might say that B's account is too sensible to be persuasive. With respect to a limited practical aim, say that of curing a patient suffering from a disease, it is a perfectly satisfactory answer to an inquiry about the relevance of a given means that it is more *likely* than not to effect the desired end. Now B's answer has quite the air of responding to such a limited practical inquiry. (One recalls Moore pointing to a hand to dispel a philosophical doubt about the existence of material objects.) But once the rationale of the entire inductive practice has been brought into question and made to seem problematic, in a fashion that is sufficiently familiar after centuries of debate about the 'justification of induction', any *use* of the key terms 'good reason' and 'likelihood' will be apt to seem question-begging or circular.

We can imagine our questioner elaborating his perplexity as follows:

A. Let me grant, at least for the sake of argument, that the way in which we talk about reasons and likelihood does require me, on pain of violating ordinary language, to call *P* a good reason for *K*. But why should we—why should I—talk in that way? If you *use* the crucial terms in your defence, you are begging the very question at issue. A lunatic who used the term 'good reason' in what would seem to us a crazily improper way, might be able to *prove* to us that, in accordance with *his* use of 'good reason', he had good reason to use the phrase in the way he does. But that would not show him to be right. Why, then, should your answer be any better?

I believe that a challenge of this sort cannot be brushed aside as absurd on the face of it. To retort: 'We talk the way that we talk and have the concepts that we have—and that's the end of it' must seem insufferably dogmatic to anybody of a genuinely philosophical turn of mind. Why *should* we talk as we do? Why can't we change the way we talk, if that should seem desirable? What is there to show that, after all, reliance upon inductive arguments is better than mere superstition?

Instead of meeting the challenge head on, I propose first to see whether any connection can be established between the procedural purpose of our inference and the *general* purpose of the activity that it instantiates. For we must bear in mind the possibility that what underlies the ordinary language link between 'good reason' and 'likelihood' is a more intelligible connection between sound reasoning 'on the whole' and the general aims of inductive reasoning. I shall need, first, to recall some familiar points.

6 *The generality of inductive reasons*

Philosophers have often commented on the 'implicit generality' of reasons. To offer a reason, whether for a belief, an action, a moral verdict, or something else, is necessarily to rely upon some real or fancied character of the situation in question. To imply that this character is relevant to whatever is being defended is necessarily to hold that some general connection obtains between the invoked character and something else. If I say that the risk of fire is a reason for not smoking in bed, I imply that the relevant character, smoking in bed, is generally connected with a risk of fire. The relevant generalisation need not be the unqualified one, *Whenever anybody smokes in bed, there is risk of fire* (although that is roughly true and relevant) but may be the more guarded one, *Whenever anybody smokes in bed in such and such circumstances, there is risk of fire.* But unless one has some such generalisation in reserve, talk about a 'reason' is improper. It follows that what is claimed to be a reason in one case is necessarily put forward as something that *would be* a reason in a class of similar cases, where the definition of that class depends upon what is taken to be relevant in the case in hand. The reference to what *would be* the case is essential—in other words, the supporting generalisation must be 'lawlike' and not merely 'accidental'.[1] If I say, to vary the example,

---

[1] A valuable feature of Peirce's last discussions of probability is his frequent insistence upon the importance of bearing in mind what *would be* the case and not merely what *will be* so. For example: 'I am, then, to define the meanings of the statement that the *probability*, that if a die be thrown from a dice box it will turn up a number divisible by three, is one-third. The statement means that the die has a certain 'would-be'; and to say that a die has a 'would-be' is to say that it has a property,

'His inflicting unnecessary cruelty in vivisecting the cat is a reason for morally censuring him', I must imply that infliction of unnecessary cruelty to animals by anybody *would*, so far, be a reason for moral censure. There would be a patent absurdity in saying, for instance, 'This act in which somebody inflicted unnecessary pain is morally censurable, but some cases, differing in no relevant respect, *would not be*'.

Let us apply these elementary considerations to our specimen inductive argument. I said that belief in the truth of the amalgamated premise, *Most businessmen are Conservatives and Grumbacker is a businessman* is a good reason for belief in the truth of *Grumbacker is a Conservative.* Here, the class of cases similar in all relevant respects is easy to discern: In calling the argument sound, I am committed to holding that belief in the truth of any premise of the form *Most B are C and A is a B* would always be a good reason, so far, for belief in the truth of a conclusion of the form *A is a C.* And, conversely, if the general statement about a good reason is true, its application to the case in hand necessarily follows.

This offers some hope of connecting the general purpose of what I have called the inductive activity with the immediate purposes of the particular inductive performances instantiating that activity. We may hope to pass from what *would be* a good reason in each of a class of cases of inductive inference to what *is* a good reason in any given case. Let us see how this might be attempted.

7 *A deductive link between general inductive purpose and immediate inductive purpose*

Consider the following example. I draw a ticket from a lottery of a thousand tickets, knowing that just one of those tickets will earn a

---

quite analogous to any *habit* that a man might have' (*Collected Papers*, 2.664, p. 409). Also: 'A probable deductive argument is valid, if the conclusions of precisely such arguments (from true premises) *would* be true, in the long run, in a proportion of times equal to the probability which this argument assigns to its conclusion; for that is all that is pretended' (ibid. 2.781, italics added). On the other hand, Peirce often slipped into talking, mistakenly, in my judgment, about what *will* be the case: 'The validity of a presumptive adoption of a hypothesis for examination consists in this, that the hypothesis . . . is selected according to a method which *must* ultimately lead to the discovery of the truth, so far as the truth is capable of being discovered, with an indefinite approximation to accuracy' (ibid.). Peirce's reasons for thinking that the truth *must* ultimately be approximated to by inductive methods seem to me incorrect.

prize. I argue as follows: 'Nearly all the tickets will fail to win a prize; therefore (almost certainly), my ticket will fail to win a prize.' If the soundness of this argument is challenged, I can defend it as follows: 'Each and every one of the other 999 persons who draw a ticket will be in the same cognitive situation as I am. What counts as a good reason for me must equally count as a good reason for each of them. Now if all the thousand ticket-holders were to argue as I did, it necessarily follows that 999 would affirm a true conclusion and only one would (though with good reason) affirm a false one. Thus the activity is certain to lead to true conclusions *on the whole*, which is more than can be said, with certainty, of any other procedure for reaching the desired conclusion. Conversely, if the general purpose of the activity is taken to be that of getting true conclusions on the whole, it follows that *each* of the ticket holders has a good reason for arguing as I originally did, and for not arguing to the contrary conclusion. What is a good reason for all is necessarily a good reason for each.'

This defence seems to be sufficient if the general purpose is acceptable. Of course, if a prospective buyer of a lottery ticket were to say that he had no interest whatever in what would happen to the other ticket-holders, but wished only to know the fate of his own ticket, the defence I have suggested would seem pointless to him. But then he would be unable to consider *reasons* for having one belief about the outcome rather than another. To weigh inductive reasons is necessarily to be in a posture of considering what *would* happen in a class of relevantly similar cases—not, as Peirce wrongly suggested, because the individual identifies his interest with that of the unbounded community,[1] but rather because it is of the essence of reasons to be 'implicitly general'.

I wish now to urge that the defence I have offered for my conclusion about the lottery ticket holds good whether or not the remaining tickets are in fact all drawn. Even if some of the tickets remain unsold and undrawn, we can still argue as I did before; it will still be true that 999 reasoners drawing the parallel conclusion *would be* right. For, to be right 'on the whole' can only mean, for the reasons I have already sketched, to be such as to yield a majority of true answers in a set of *would-be*'s. Take the extreme case in which I draw a ticket, on the terms already stated, knowing in advance that all the other 999 tickets will immediately be destroyed. Even then, appeal to what would happen 'on the whole' is as relevant and as conclusive as before.

---

[1] 'It seems to me that we are driven to this, that logicality inexorably requires that our interests shall *not* be limited. They must not stop at our own fate, but must embrace the whole community. This community, again, must not be limited, but must extend to all races of beings with whom we can come into immediate or mediate intellectual relation. It must reach, however vaguely, beyond this geological epoch, beyond all bounds. He who would not sacrifice his own soul to save the whole world, is, as it seems to me, illogical in all his inferences, collectively. Logic is rooted in the social principle' (*Collected Papers*, 2.654).

In some such way as the foregoing, connection can be demonstrated between the immediate procedural purpose of using a statistical syllogism and the general purpose of the activity instantiated by such a performance.

I do not think I need to discuss the rationality of the general purpose of being right—of getting a true conclusion—'on the whole'. If we desire truth about the unknown, but cannot hope to get it each time, we must be satisfied with a method that would serve most of the time. But the line of defence adopted will hardly suffice, without important modification, to justify other modes of inductive argument that have an equal title to be regarded as basic.[1] For instance, it is hard to see how the defence can be adapted to fit a case where the reference class is indefinite or potentially infinite in membership (as when a ball is drawn from an urn with replacement); and it is even harder to see what sense could be made of the notion of 'being right on the whole' in arguments whose conclusions are general.[2]

In order to illustrate further some of the problems that can arise,

---

[1] One of the most important of these is the argument from the composition of a sample to the approximate composition of the population from which the sample is drawn. Suppose m/n members of the population have a certain character C. Then, most samples containing n members will have a frequency of occurrence of C falling within the range $m/n+\delta$, for suitable choice of δ. This would allow us to argue that *most* samples will approximately match the population and so to construct a defence somewhat like that already offered. Cf. D. C. Williams, *The Ground of Induction*, for a similar argument.

[2] It is worth recalling that Peirce did not rest the 'validity' of inductions upon success 'on the whole'. On the contrary, he said, 'The validity of induction is entirely different; for it is by no means certain that the conclusion actually drawn in a given case would turn out true in the majority of cases where precisely such a method was followed' (*Collected Papers*, 2.781, p. 501). And in the same vein: 'It is nonsense to talk of the probability of a law, as if we could pick universes out of a grab-bag and find in what proportion of them the law held good' (ibid. 2.780, p. 500).

I shall now consider the type of inductive inference sometimes called *eduction* (inference from particulars to particulars) and will try to discover how far one could get in establishing in this type of case a deductive connection between immediate and general purpose.

8 *A quasi-pragmatic vindication of proportional eduction*

By a *proportional eduction* I mean an inductive argument of the following form: *Most of the observed A's have been B; therefore (probably), The next A to be observed will be B.* My aim, as it was in the case of the statistical syllogism, is to see how far one can hope to go in establishing some necessary connection between the 'immediate purpose' and the 'general purpose'. Since I see no prospect of finding a line of argument closely analogous to the one I have already used, I propose instead to follow a somewhat different tack.

A slight change of perspective will be helpful. For we can, without substantially changing our task, now conceive of the principle of the proportional eduction as governed by a certain *policy*, rather than by the corresponding principle. We can think of any particular eduction as exemplifying and conforming to the policy, 'When most observed *A's* have been *B*, expect the next *A* to be *B*, unless there are sufficiently strong countervailing reasons to think otherwise'. I want to see what can be said for this policy in terms of success 'on the whole'. In order to do so, I shall now describe the following, highly artificial, situation in which the policy might be applied.

The finite prediction task

Imagine that you are to be shown, one by one, a hundred objects, each having one or other of two qualities (call them 1 and 0, respectively). Before each object is shown, you are to predict its character (i.e. whether it will be 1 or 0), whereupon you will at once be told the right answer. You are allowed to keep a cumulative record of your predictions and of the right answers. The object of the exercise is to get the highest number of correct answers by the end of the test run. I shall also stipulate that no further information is available: you are completely in the dark about the nature of the two characters here called 1 and 0, and you have absolutely no information about whether the series of 1s and 0s will be a random one, whether the experimenter who supplies the objects will make his choice depend upon your own performance—in short you are and will remain until the end of the test run completely uninformed about the design of the series and the plan, if any, behind its generation. Given all this, is it possible to have a rational policy for performing the set task?[1]

Some men of good judgment will say that in the situation envisaged we are so close to the condition of 'complete nescience' envisaged by certain writers on induction as the setting for ultimate justification of inductive procedures[2] that no rational policy is defensible and blind guessing would be as reasonable as anything else. I shall however try to defend the following policy (call it *P*): (i) Whenever the observed segment of the run shows a predominance of one character over the other (more 1s than 0s, or vice versa), to predict the immediate occurrence of that character; (ii) whenever the 1s and 0s have occurred equally often in the observed segment, to predict 1 and 0 alternately (i.e. predict 1 the first time this happens, 0 the next time, and so on);[3] (iii) to scrutinise the observed segment constantly for evidence of order, and to modify the application of the first two clauses accordingly (I shall explain the meaning of this very soon).

The first clause in this statement of the policy in effect prescribes the use of proportional eduction whenever it is applicable; the second is merely an arbitrary way of ensuring a determinate prediction when the proportional eduction fails; and the last is a somewhat vague injunction to override the eductions specified, whenever sufficient evidence of regularity appears in the series. Vague as it is, the last clause seems essential. To conform solely to the first two clauses when faced by a series that continued to deliver a 1 and an 0 alternately would, plainly, be highly unreasonable.

---

[1] The test situation described is essentially that used in many simple experiments on choice under uncertainty. See, for instance, William K. Estes, 'Individual Behavior in Uncertain Situations: An Interpretation in Terms of Statistical Association Theory', in R. M. Thrall (ed.), *Decision Processes* (New York, 1954), pp. 127-137.

'In these experiments the outcome of the situation has no utility for the individual (except that of being right or wrong in his guess) and the information available to the individual is restricted to what he can obtain from observing a series of replications of the situation.... The subject is given no information about the conditions of the experiment except that one of the E_j [the events to be predicted] will follow the signal S on each trial; he is instructed to do his best to make a good score (in terms of correct predictions), and to make a prediction on each trial regardless of how uncertain he may feel about the outcome' (op. cit. p. 128).

[2] 'Simple induction ... is the first step in the emergence from total nescience' (R. Harrod, *Foundations of Inductive Logic*, London, 1956, p. 67). He seeks a defence of inverse probability 'as a method of rescuing us from initial nescience' (op. cit. p. 87), construed as a state where, notably, we cannot employ a Bayesian resort to *a priori* probabilities of the hypotheses under test.

[3] Or use *any* device that will ensure a definite decision. The purpose of this clause is merely to avoid any indeterminacy in applying the policy.

Let us put aside, for the moment, complications arising from the presence of the last clause in the statement of the recommended policy, by supposing that we are dealing, for the time being, only with series in which the initial segments manifest no evidence of any special order (series which are random in the distribution of 1s and 0s, although with a possible predominance of the one character over the other). We can then argue in defence of P in the following way: Suppose one character, say the 1, predominates in the entire test run of 100 items; then in that run there will be a certain number, n, by which the total number of 1s exceeds the total number of 0s. Let us call $n-1$ the *security*. Then it is a fact, though one that is not quite obvious, that by following P, we can be certain, in advance, of ending with a final score of successes *equal to or greater than the 'security'*. Furthermore, any deviation from the recommended policy runs the risk of reducing the final score below that guaranteed lower bound. P is what game-theorists call a 'maximin' policy, since the worst that can happen by following it is no worse than the worst that could happen upon adopting some other policy.

It is not hard to see why the final score obtained by following the recommended policy, P, must be equal to or greater than what I have called the 'security'. Let us call the character (1 or 0) that has occurred more frequently at a given point in the test run, the *leader*, and let us call predicting the immediate recurrence of that character, *backing the leader*. Suppose 1 ends up as the ultimate leader—the *winner*, let us say—by a margin of n more total occurrences than 0 has. Then, obviously, 1 must have been the leader at least n times: it needs to have been ahead at least n times in order to build up its final lead (more often than n times if 0 is ever ahead). At the outset we do not know which character will end as the winner, but by always backing whichever character happens to be in the lead at a given point we can be *sure* of ending on the right side.

A helpful way of looking at the matter is this. Since we want to be right in our predictions as often as possible, we should like to know whether most of the items to be shown to us are 1s or whether most of them are 0s. Had we been told at the outset that most of the items were 1s, say, the rational policy would clearly have been to back a 1 all the way, for this would have guaranteed a final score equal to the total number of occurrences of the more frequent character. By the policy of constantly backing the leader, we do end by backing the winner—the character that occurs the more frequently—and meanwhile we have been backing it at all times when it was ahead, without knowing that we *were* backing the winner. So the recommended policy for the task can also be thought of as a way of approximating to the truth about the total run—or at least to the truth about which character finally predominates.

The suggested defence of policy *P* can now be set out as follows: For each given test run, *R*, there is an associated but unknown number, *s*, greater or equal to zero, i.e. the number that we have called the ' security '. If *s* is equal to zero, we have nothing to lose by following *P*, for, on the special assumption made, that no regularity can be detected in the successively revealed initial segments of *R*, *R* will be random and no policy is any better than any other. On the other hand, if *s* should be greater than zero, which it may be, we would have the advantage, by following *P*, of having a built-in guarantee against ending with less than *s* as our final score. Since no other policy has this advantage, *P* is preferable to any competitor.[1]

It must be admitted that this is, at best, a very weak defence. It does, to be sure, establish *some* necessary connection between the proximate procedural purpose of drawing the predictive conclusion in a certain stipulated way and the more remote general purpose of doing well—or as well as can reasonably be hoped for—' on the whole '. It guarantees the modest prospect of insurance against a very low score *in a particular case*—but is that enough to satisfy us?

Here I am, with a particular test run before me and the task of getting a true eductive conclusion as many times as I can in my 100 tries; why should I follow *P*? Well, if the series manifests sufficient regularity for me to detect it, *P* may well lead me astray (as we shall see very soon); and if the series is irregular, with equal occurrences of the two characters, *P* is no worse than any other policy—i.e. to be candid, no good at all; but if P is thoroughly irregular, with a predominance of one character over the other, then following P will make it certain that my score will not fall below a certain, unknown, number s. This defence of P is so hedged about with provisos as to be less than persuasive.

---

[1] It might be noticed that *P* is a very conservative policy that cannot at best produce a very high score. Let t be the total number of occurrences of 1 (the character assumed to predominate). If the arrangement of the terms of the series is sufficiently favourable we might get an assured score of $t-1$, but no more (apart from lucky guesses at points where neither character is in the lead). Let e be the number of points at which the 1s and 0s have occurred equally often; then, all that we could count upon is $t-e-1$, which cannot be less than s, but need not be more.

Would it be a good policy to follow P if I had to take the experiment a large number of times? There is no argument to show that it would. The recommended efficacy of the policy depends upon the same procedure being applied, without interruption, from the very start of the run. Now, if I am required to repeat the experiment on another occasion, I cannot consider the next test run as an extension of the old one, because I presumably have no record of that earlier one. The most one can say is that if the policy was rational in one case, it is rational in every similar case.

A serious weakness of this policy is that it considers only possible losses, without balancing them against possible gains. What is at stake can be brought out by contrasting P with a modified policy, which I will call P′. The modified policy is simply that of following P for the first fifty terms and then starting afresh for the last fifty predictions, as if they constituted a new test run. Let the 'securities' of the two halves of the test run,. considered separately, be s_1 and s_2 respectively (and let s as before be the security for the whole run). Then it is easy to see that if the same character predominates in the two halves, we must have

$$s_1+s_2 = s-1$$

That is to say, by performing our task in two independent instalments, we risk losing one point in the overall guarantee. (The first member of the second half of the total run, where we can only guess, might have been a place where we ought to have been backing a 1.) But now suppose that the two halves show *different* predominating characters, say with 1s predominating in the first half, 0s in the second, and 1s predominating over all. Then the equation between the securities will be

$$s_1-s_2 = s-1$$

Our total guarantee in the interrupted procedure, P', viz. s_1+s_2, will then be greater than the security, s, obtainable on P. In short, we stand to gain something that may be substantial, at the risk of a small loss (one point off the total security) if the two halves are not homogenous with respect to the predominance of one and the same character. By using *P* rather than *P′* we are, as it were, counting on the homogeneity of the series; by preferring *P′* to *P* we would be counting on the heterogeneity of the series. (Similar considerations would apply to a proposal to use another policy, say *P″*, involving separate consideration of three successive partial segments of the test run.)

In the absence of any consideration of probabilities (and perhaps associated utilities of success and failure) there seems to be no rational way of deciding between *P* and *P′*—or indeed between *P* and many other policies. One might say the choice depends upon the subject's relative interest in winning or losing. If he is more concerned to avoid loss than to acquire gain, he should prefer *P*, but then he runs a risk of failing to pocket a considerable gain. If, however, he would rather take a slightly increased risk of loss for the sake of having a chance of great gain, he might well prefer *P′*.[1]

Let us now return to what we have so far excluded, the use of the third clause of the policy, which was, it will be remembered, to check each observed segment for possible evidence of regularity. Suppose we have had the following initial segment: 1010101. Adherence to the policy *P* would require the subject to predict a 1, but if he thinks that there is now good evidence for a regularity in the series he might well prefer to predict an 0. Here, there is a sharp conflict, which I can see no way to resolve without recourse to questions of probability (and perhaps utility). To abandon *P* at this point is certainly to diminish the 'security'; but perhaps that is worthwhile, in view of the gain to be expected if the hypothesis of regularity is verified.

Is the chance of gain on the hypothesis of regularity worth the reduction of security involved in abandoning the conservative maximin policy? I see no way of answering this question in the absence of further information. In practice, I imagine, anybody actually faced with the task I have described would decide, sooner or later, that a long initial segment of alternating 1s and 0s justified him in betting on its continuation. Called upon to rationalise this, one would probably argue that the conservative policy P is tailored to the case of random distribution (for it is only then that its safeguards are least in doubt), while a *long* run of alternating 1s and 0s makes the hypothesis of randomness progressively less likely. This argument is, clearly, an exercise in inverse probability, or in the rational choice of hypotheses—at any rate, is based upon considerations a great deal more complicated than anything that I am here seeking to defend.

---

[1] In experiments of the kind cited in footnote 1, p. 193, groups of subjects, considered as a whole, tend to approximate asymptotically to a strategy which 'matches' the frequency in which the characters to be predicted occur in the (random) series of their presentations. (Thus, in my own statement of the conditions, if 1s appear twice as often as 0s, without any noticeable regularities in occurrence, the average behaviour of a large group of subjects approximates to predicting 1 twice as often as 0.) This is clearly *not* the policy I have been recommending, which is approximately the same as the 'pure strategy' which would be recommended by game-theoretical considerations.

9 *The moral of the preceding exercise*

The outcome of our attempted quasi-pragmatic vindication of proportional eduction is disappointing, for the connection between the rule governing the performance of eduction and the presumptive aims of the activity instantiated is tenuous to the point of non-existence. The chief lesson to be drawn from the exercise is perhaps the implausibility of hoping to establish a necessary connection, even of so tenuous a sort, between the purpose of the performance and the purpose of the associated activity, so nakedly conceived. It would hardly be a parody on the defence we have attempted to compare it to an effort to justify the wearing of a hat by arguing that if we were to wear a hat in a state of otherwise complete nudity, we might have a better chance of keeping at least our head warm than if we went bareheaded. A state of almost total ignorance concerning the relevant empirical situation is about as rare, and needs as much special contrivance, as a state of complete nudity. Were we to be actually faced with the kind of predictive task I have envisaged, we would, I believe, try to assimilate it to a condition of substantial antecedent knowledge. Thus, it would certainly be rational to speculate concerning the motives of the experimenter, what plan he might be using in generating the terms of the series, and so on—in short, to formulate hypotheses, however ill-supported by evidence, as we should normally do in any authentic empirical investigation. In the absence of any access to relevant information about the background of the set task, our inductive procedures will be, at best, pallid and ineffectual efforts to behave as if we had a solid basis for full-blooded inductive inference.

If challenged to defend 'eduction' as an isolated procedure we might, of course, fall back upon the linguistic argument I used before. It is, I think, correct to say that all the observed A's having been seen to be B is *so far* a good reason for thinking that the next A will (probably) be a *B*, and to appeal to the way we use the key terms in justification. But that gets us a very little way, at best, and leaves us exposed to the old accusation of circularity.

10 *Links with the entire practice?*

The reason for our setback may, of course, be simply lack of competence. It is notoriously hard to establish a negative, and correspondingly hard to show that there is *no* necessary connection between success in performing a 'proportional eduction' correctly and some intelligible general purpose of the activity instantiated. Instead of what I called the 'security', it may be possible to find some other feature of a long series of applications of the type of inductive argument in question that would supply the missing link between the short-term and the long-term purpose. I am strongly inclined to doubt this, however, if only because so many talented philosophers have tried it without success.[1] I would therefore like to raise now the possibility that there is something wrong with the approach adopted. Perhaps it was a mistake to seek to isolate the activity of 'proportional eduction' from the wider practice to which it belongs (as it would be to seek to establish the utility of addition, without reference to the function of the whole of arithmetic). Perhaps the place to find a connection between induction and intelligible human interests in arriving at the truth is in the entire *practice* and not in some artificially dissected component of it.

Well, what is the overall purpose of the inductive practice? Can it be anything but arriving at an approximation to the truth? And is not the past experience of the race some evidence of its success in achieving that purpose? Certainly there is a strong inclination, on the part of the sophisticated no less than the naïve, to point to the past successes of the whole inductive practice as some kind of justification. Surely, one is inclined to say, the fact that the inductive practice, broadly conceived, has worked so well in the past *must be* a good reason for continuing to trust it.[2] Well, much depends upon how the inductive practice and its global purpose are conceived—whether as revelation of the most general and systematic truth about the universe, or as a series of interconnected assertions having high confirmation, or as a progressively 'self-correcting' procedure for approximating in the infinitely long run to the truth (Peirce) or, as a rigorous and sophisticated strategy for eliminating plausible and interesting errors (Popper). Whatever answer is given, there will be many who will be quick to level the charge of circularity against *any* such attempt. Suppose the global purpose of inductive inference be isolated as Q (Q being such as to make the more immediate purposes of particular inductive activities intelligible in their relation to it) and suppose also that there is no dispute as to the practice having sufficiently achieved Q in the past. Consider, now, an argument having the form: *I has achieved Q in the past; therefore (probably), I will continue to achieve Q in the future.* Is there any reprehensible circularity in an argument of this form?

---

[1] See, in this connection, especially D. C. Williams, *The Ground of Induction* (Cambridge, Mass., 1947) and Roy F. Harrod, *The Foundations of Inductive Logic* (New York, 1956).

[2] Cf. the following answer to the question whether the Laws of Nature will continue to hold: 'I reply that, unless the constitution of the Universe shall be changed to an extent which I cannot now even conceive, they will so continue, and that no

I believe the following assertions concerning this kind of argument are correct:

1. The argument is *sound.* That is to say, anybody who believes in the truth of the premise would be justified on inductive grounds in believing in the truth of the conclusion and, consequently, in asserting that conclusion.

2. We have, for the reason set out in the premise of the argument, strong grounds for holding that the inductive practice will continue as it has done in the past to serve its distinctive purposes.

3. There is no *formal circularity* in the argument, i.e. nobody using the argument needs to assume *as a premise* that the inductive practice

---

reasonable man has any practical doubt as to their continuance. And why? *Because they are confirmed by the whole of our own experiences,* which . . . is of enormous extent and variety, *by the experiences of our ancestors, and by all that we can ascertain of the past history of nature* [he is speaking of the law of gravitation and the proposition that animal and vegetable life cannot exist without moisture], while their reversal would involve the reversal of almost all the other laws with which we are acquainted' (Thomas Fowler, *The Elements of Inductive Logic,* 6th edn., Oxford, 1904, p. xi, italics added). In the tradition of Mill, Fowler makes this belief in the 'continuance' of laws of nature depend upon the 'hypothesis' that 'the same causes will continue to produce the same effects'. He concludes: 'What would happen if this expectation were ever frustrated, it is absolutely impossible for us to say, so completely is it assumed in all our plans and reasonings' (ibid.).

will continue to work successfully in the future—or anything from which this would follow.

4. Nor is there any more subtle circularity such as would be involved in anybody's necessarily having to believe (or even to know) the conclusion to be true before he could be justified in presenting the argument as sound.[1]

What, then, does the accusation of circularity, commonly brought against any argument of this form, amount to? The accusation reduces to something like the following, I believe: ' In offering the inductive argument in support of the continuing efficacy of the inductive practice, you must be at least believing (if not actually assuming) that your argument is sound. But then you must be believing (or actually assuming) that a certain mode of inductive activity (simple eduction) *will* continue to work in the future. Now this is part of what is in question here. For unless simple eduction works, the entire practice cannot continue to be efficacious. You are, in part at least, assuming the very thing to be proved.'

The following rejoinders can be made. It must be conceded to the critic that anybody offering the inductive defence of the inductive practice in good faith must hold (as he undoubtedly will) that the form of inductive argument he is using is sound. But, on the type of view here being outlined, that is not the same thing as holding that the argument 'works' in the sense of being right ' on the whole '—indeed we have had serious reasons to doubt whether *that* view can be defended. The most that can be said is that the type of inductive activity called ' eduction ' is conducive to the aims of the overall activity—and that is guaranteed, not by any contingent empirical considerations, but by an *a priori* connection between that particular activity and the entire practice. What we *mean* by the ' inductive practice ' includes eduction; and what we mean by ' more likely than not ' guarantees that the conclusion of an eduction will be more likely to be true than the contrary conclusion. The critic who raises the objection of circularity is really objecting to the use of inductive language and the concepts expressed by such language. He is objecting to the use of typically inductive concepts in advance of a demonstration that will justify such use. He is, it might be added, in a radically incoherent position. In using, as all of us do, language that purports to refer to persisting physical objects, he is himself committed to inductive beliefs at least as sweeping as those he wishes us to forego. But an argument *ad hominem* will accomplish little at this juncture.

---

[1] I have argued in the past, more narrowly, for the non-circularity of inductive support of an inductive rule. See my controversy about this with Peter Achinstein in *Analysis*, 1961-62, **22**, 138-141; 1962-63, **23**, 43-44 and 123-127. Also Henry Kyburg's comment on the dispute in *American Philosophical Quarterly*, 1964, **1**, 260-261.

We may grant that, without the use of inductive concepts, no inductive defence of the inductive institution can get off the ground. It is worth pointing out that in this respect the defence of induction is in no worse case than the defence of deduction. For after all what is to be said by way of justifying deduction? Is it not, essentially, that if the premises of a valid deductive argument are true, and if the argument is valid then, necessarily, the conclusion is true (with an implicit appeal to our interest in deriving new truths from old). But how is *this* established? How else except by appeal to the ways in which we do actually use the key terms—or, if preferred, by the use of deductive concepts of validity and the like, which presuppose the correctness of certain deductive inferences? If the alleged circularity obtains in the case of induction, it must be held to obtain equally in the defence of deduction that we commonly advance without any qualms. If our trust in induction is held to be dogmatic, the same must be said for our confidence in deduction.

If objection to the use of inductive language and inductive concepts is indeed the burden of the sceptic's complaint, scepticism concerning induction is introduced by the very terms of the demand. Given that induction differs from deduction by its very definition, a deductive proof of its efficacy is impossible in principle. If the demand is also made that no *inductive* concepts be used in justification, then an inductive defence will also be ruled out *ab initio*. But then the allegedly sceptical conclusions will spring from the arbitrary blinkers imposed by the philosophical critic. Once it has been understood that the demand for 'justification of induction' has been so constructed that it can be seen *a priori* to be impossible of fulfilment, the sceptical conclusions drawn from the impossibility of fulfilling an impossible task will seem somewhat less than compelling. If a Zen master imposes upon me the task of jumping without moving a limb, the impossibility of satisfying *him* will reveal only something about the taskmaster and could be regarded as a symptom of human incapacity only by a very muddled thinker. Now the same can be said, *mutatis mutandis*, about the alleged impossibility of solving the philosophical 'problem of induction'. If we seem unable to make the 'inductive leap ', that is because philosophers, excessively influenced by the deductive conception of inference, insist upon defining the task in a way that renders it logically impossible to jump. But if we ignore the arbitrary distortions implicit in this persuasive manoeuvre, we shall find that we can jump perfectly well. *Solvitur saltando.*

In order to forestall possible misapprehension, I would like to add that I am not proposing the inductive argument from the past successes of the inductive practice as a solution to the philosophical problem of the justification of induction. (This would be as misconceived, in my opinion, as trying to deal with somebody who thought he could see snakes everywhere by arguing that at least they don't bite.) The roots of the philosophical problem are to be found in the imposition of arbitrary and inappropriate standards of rationality. There is no way to cope with the ' problem ' that, in my opinion, offers any prospect of satisfying those to whom its solution seems necessary except by patiently exposing the underlying confusions until the alleged problem withers away. I have been trying to draw attention to a much less exciting point. I have thought it proper to ask, as we might of any human practice or institution, whether there is reason to think that it has fulfilled and will continue to fulfil the human purposes it is presumably intended to serve. My answer is the commonplace one that there is indeed good inductive evidence for thinking that our universe is of such a character that continued trust in the inductive practice is reasonable. That no stultifying circularity is involved in such a claim, and on such grounds, is a point of considerable importance —if only because so many wise men, steeped in the subject, still think otherwise. A fuller and more adequate defence against the charge of circularity would need more space than I here have at my disposal.

11 *Retrospect*

I have here been adopting a somewhat heterodox conception of inductive argument that takes it to be subject to distinctive rules of sound procedure, not reducible to the canons of deductive argument. I have further assumed that at least *some* conclusions of good deductive arguments can be ' detached ' so that one can, in favourable instances, proceed from the assertion of true premises to the justified assertion of other propositions, sufficiently supported, though not entailed by, those premises. Throughout, I have chosen to think of induction as a public practice, embodying distinctive rule-governed activities, each of which is instantiated in indefinitely many inductive performances, actual or possible. Conceiving, as seems reasonable, that the practice as a whole is concerned with the rational acquisition of truth, it has seemed plausible to expect links between the 'immediate purpose' or goal of the individual inductive performance and the more general purpose of the activity instantiated—and between the latter and the overall purpose of the supporting practice. Such links are, in a way, easily found. It is an *a priori* truth that inductive arguments of given forms are *more likely than not* to lead from true premises to true conclusions based upon those premises. Thus, the performances are weakly conducive to the general purpose of arriving at the truth. It is tempting to hope for a stronger connection than this—to establish that certain cognitive benefits *must* ensue from the use of inductive methods. Glimpses of such connections are indeed to be captured: We can see, for instance, that the 'proportional syllogism' is *bound* to yield true answers from true premises in most of the cases of its application. And even in the more problematic instance of 'simple eduction', I showed that some case can be made for an associated guarantee or 'security' against an excessive number of predictive failures. But it had to be admitted that the outcome of such argument was extremely meagre, which suggested, in turn, that a basic fallacy of method might be involved in seeking to detach particular inductive methods from the global practice to which they belong. I argued that the practice, considered as whole, could be shown, by a sound and non-circular inductive argument, to be as deserving of our rational confidence as in our unreflective moments we hold it to be. I also urged that such commonsensical reflections upon the utility and relevance of the inductive practice held little promise of being able to cope with the philosophical problems associated with induction. Having deeper roots, those stubborn perplexities demand a more fundamental therapy than the appeal to common sense.

Cornell University
Ithaca, N.Y., U.S.A.

Brit. J. Phil. Sci. (1966), Volume 17, Number 3, 205–218
Printed in Great Britain

A NEW LOOK AT THE PROBLEM OF INNATE IDEAS*

NICHOLAS RESCHER

1 *Introduction*

THE mention of innate ideas evokes little responsive sympathy in modern philosophers. When, for example, I inquired of Nelson Goodman regarding publication of a recent symposium on the topic on which he had participated, he wrote in reply that: 'Personally, I am rather of two minds about seeing this published since I feel in some ways that the less attention given the matter of innate ideas the better' (Letter of 22 July 1965). Nevertheless, I cannot help feeling that here as elsewhere there is truth in the maxim that *Where there's smoke there's fire.* A position which has in the history of the subject been held by philosophers of the very first calibre is not likely to be devoid of merit, and is certain not to be devoid of interest.

The problem we have set ourselves here can thus be stated simply: What good reasons (if any) can be found for holding a theory of innate ideas? However, before dealing with this question, we must first confront the obviously preliminary query: Exactly what are the concepts and the theses that are involved here—just what is at issue in a theory of 'innate ideas'?

2 *The Notion of an 'Innate Idea': Some Contrasts*

When we speak of 'innate ideas', two centrally important contrasts must be borne in mind:

First, the term 'innate' involves (1) a contrast with *importation* from without, the idea thus characterised being held not to be learned or in any way acquired; and there is also (2) a contrast with internal *manufacture*, the idea in question being claimed not to be synthesised or in any way generated by oneself.

Second, the term 'idea' itself involves a whole series of significant

* Received 21.x.65

contrasts. It indicates that we have to do with *concepts* or *notions*, and not, strictly speaking, (1) *knowledge* or (2) intellectual *capacities* or (3) mental *propensities*. Brief consideration of these last three items is indicated.

Innate knowledge. That there are ' innate truths '—items of information or knowledge whose *rationes cognoscendi* are in a significant way independent of experience—is a position upheld in Plato's *Meno* and supported by various subsequent philosophers. This position undoubtedly leads naturally into a theory of innate ideas. It is plausible —though not strictly necessary—to maintain that when a truth is innate the ideas with which it deals cannot be a matter of importation (acquisition) or manufacture (generation by oneself). But the contrary implication does not hold. A theory of innate ideas could be upheld without any concomitant doctrine of innate knowledge. The two positions are thus distinguishable.

Innate intellectual capacities. These relate to the whole range of man's mental abilities and capabilities. A great many intellectual capacities are unquestionably acquired—the ability to solve calculus problems or to speak French, for example. Others can with legitimacy be described as innate—the ability *to learn* the calculus or *to attain fluency* in French, for example. For these are capacities which, in the very nature of the concepts at issue, cannot unqualifiedly be characterised as acquired rather than present *ab initio*. All of those intellectual capacities (like that for language-learning) in virtue of which we are able to acquire other capacities (like that of conversing in English), are in a basic, if quite unexciting sense to be described as innate. Capacities of this sort appear to be built into our very conception of man as a creature capable of ' human reason '. Nothing mysterious—and, alas, but little of epistemological interest—is at issue here.

Innate mental propensities. Here we have to do with the range of tendencies, inclinations, and dispositions relating to the functioning of the human mind. These could presumably be of (among other things) a *propositional* orientation—i.e. be a matter of tendencies to believe certain theses. Or, and from the standpoint of innate ideas more relevantly, they could be of a *conceptual* orientation—i.e. be a matter of tendencies to conceive, that is, to conceptualise certain things in some one particular manner. The espousal of a theory of innate ideas would thus compel one to accept innate mental propensities (although the reverse implication does not necessarily hold).

3 *Historical Considerations*

To bring into sharper focus the themes that are at issue in the discussion of innate ideas it is useful to examine briefly some facets of their history.[1]

Descartes. The doctrine of innate ideas came to prominence in modern philosophy in the system of Descartes. His attention focused principally upon the idea of God, arguing (in Part III of the *Meditations*) that this idea is of such a kind that it (1) cannot be presented to the mind on the basis of sensation, (2) cannot be presented to the mind on the basis of reflection (i.e. the mind's internal self-perception), and (3) cannot be constructed by the mind out of sensory and reflexive materials. 'Consequently', writes Descartes, 'the only alternative is that it [i.e. the idea of God] is innate in me, just as the idea of myself is innate in me.' Descartes thus reaches his conclusion that the idea of God is innate, not by adducing any positive evidence to this effect, but wholly by preceding along the *via negativa* of the disjunctive syllogism: We have the idea in question in the mind; it was not imported there by the usual channels; it could not have been manufactured there; *ergo* it must have been there from the first.

Locke. The brunt of Locke's attack in Book I of the *Essay Concerning Human Understanding* is directed at 'innate principles', innate truths, rather than innate ideas. (His primary target was apparently not the innate ideas of Descartes but the inbred 'self-evident principles' of the British scholastics and the Cambridge Platonists.) However, Locke leaves no doubt that a theory of 'innate notions', i.e. ideas, is equally unacceptable to him. The strongest argument for innate ideas would, on his view, be their universal recognition by all mankind. But such universal recognition would—even if actually extant—provide only a weak basis for the theory. And moreover, there is every reason to reject any claim of universal recognition.

> Children, idiots, savages, and illiterate people, being of all others the least corrupted by custom or borrowed opinions, one might reasonably imagine that in their mind there innate notions should lie open to everyone's view . . . (*Essay*, Book I, chap. ii, sect. 27).

Here, as elsewhere, Locke treats the theory of innate ideas as immediately refutable by straightforwardly empirical considerations.

---

[1] A conspectus of the history of the theory of innate ideas from the Cambridge Platonists to Kant can be found in W. Windelband's *History of Philosophy* (Part V, chap. i, sect. 33).

Leibniz. In his *New Essays Concerning Human Understanding* Leibniz in effect rejects Locke's universality-of-recognition construction of innateness and goes back to an essentially Cartesian conception of the matter. It is not requisite that everyone recognise an innate idea—it can lie latent in the mind. We may be—indeed we may *have to be*—led to an innate idea by a suitable course of experience. (Leibniz explicitly rejects the thesis that *Whatever we learn is not innate.*[1]) What renders the idea innate is a matter of content not of heuristics; the key consideration is that the idea cannot be *constructed* out of the deliverances of experience: be it external perception or internal reflection. The ' cannot ' here is of an intricate kind—neither strictly logical nor strictly factual, but *functional.* Its force is roughly this: Given the nature of the mechanisms of human experience (an essentially empirical matter) it is then logically impossible (but only *conditionally* impossible with respect to this assumed empirical basis) to construct from such elements an idea with the characteristics in question. (This is like saying—' Given the characteristics of my typewriter, it is impossible (now *physically* impossible) for me to write Chinese with it.') The force of such an essentially hypothetical ' cannot ' is logical, but is nevertheless such that it essentially involves factual considerations: Given a description of the mode of functioning, of the task at issue, and of the relevant principles of operation (i.e. natural laws), the performance of the task can be shown to be logically impossible.

Kant. One Kantian version of an innate idea is a ' pure concept ' of the understanding—a concept not derived from experience as contents thereof, but as essentially structural in representing mind-imposed preconditions for the having of experiences. Moreover, particular importance attaches to the ' Transcendental Ideas ', equal in number of the modes of relationship of which the mind is capable through the mediation of the categories of thought. They ' transcend the limits of possible experience ' in that they cannot possibly occur within experience but rather relate to its limits—they have no *constitutive* applicability to the contents of experience, but only a *regulative* applicability: ' den Verstand zu einem gewissen Ziele zu richten, in Aussicht auf welches die Richtungslinien aller seiner Regeln in einem Punkt zuzammenlaufen.'[2] The mode of analysis is not so much functional, as in Leibniz, but rather presuppositional: Given that a conscious experience of a mind is the sort of thing it empirically is, what sorts of ideas must be inevitably applicable to its deliverances? The direction of the analysis is reversed from that of the functional approach.

---

[1] *Philosophische Schriften* (ed. C. I. Gerhardt), vol. v, p. 70

[2] *Kritik d. reinen Vernunft*

Linguistic Neo-Kantianism. In Cassirer's variety of neo-Kantianism, the *a priori* concepts are located not in the human intellect as such, but in the particular linguistic apparatus it employs (and *some* form of which it must perforce employ). We are to look for imposed concept-preconditions not in the mind that assimilates experience, but in the conceptual system of the languages by means of which a given mind organises its experiences. There is no reference to a generic mechanism for 'having experience'; it is the concrete machinery available for the communicable characterisation of experiences in natural languages that provides the functional locus of 'innate'—or rather now simply *a priori* (in a language-relativistic sense)—concepts. The fundamental 'categories' inhere not in human thought as such but in the 'forms of human life' that provide the ecological matrices for the development and utilisation of natural languages:

> The ascent to higher levels of abstraction, to more general and comprehensive names and ideas, is a difficult and laborious task. The analysis of language provides us with a wealth of materials for studying the character of the mental processes which finally led to the accomplishment of this task. . . . Every classification is directed and dictated by special needs, and it is clear that these needs vary according to the different conditions of man's social and cultural life. In primitive civilisations the interest in the concrete and particular aspects of things necessarily prevails. Human speech always conforms to and is commensurate with certain forms of human life. (*An Essay on Man*, chap. viii, *ad fin.*)

Another version of Neo-Kantianism is operative in the line of thought presented in an interesting article by P. H. Nowell-Smith.[1] Taking a functional view of language as a purposive instrument for communication among rational beings:

> Since making a statement is a purposive activity, the concept of making a statement can only be understood in the light of the purposes for which statements are made; and since it is a rational activity it must be governed by rules which together constitute necessary and sufficient

---

[1] 'Contextual Implication and Ethical Theory', *Proc. Arist. Soc. Supp. Vol. XXXVI* (London, 1962), pp. 1-18

CALCUTTA CENTRAL LIBRARY

> conditions of anything's being used for these purposes. Prominent among these . . . is the communication of beliefs . . . ; so that the rules for making statements are rules to which any device must conform if it is to be capable of being used for the communication of beliefs.

Given this approach, it would seem to be a natural step—although it is one which Nowell-Smith himself does not take—to make a ' transcendental deduction ' of those concepts which, on the view advanced, are inherently underlying the use of ' any possible language '. This would presumably underwrite certain ideas as ' innate ' (in the technical sense of necessarily rather than contingently ' language-inherent '). The obvious candidates would be such ideas as those of ' truth ', ' believing-to-be true ', ' asserting-to-be true ', and perhaps such a derivative as ' lying ' (i.e. ' asserting-to-be true but believing-not-to-be true '). Moreover, the concepts of ' rule ', ' rule-conformity ', and ' rule-violation ' would also appear to be eminently qualified candidates.

4 *The Issue Reformulated*

In the light projected against the backdrop of this historical excursus, we can now reformulate the question of innate ideas in the way in which it is preposed that it be regarded here. We shall assimilate the controversial issue of innate *ideas* to the rather more inocuous issue of innate *propensities*, but propensities of a very special type—viz. propensities to arrive at and utilise certain ideas. Three sorts of propensities are specifically at issue here:

1. Propensities on the part of men to arrive—regardless of environmental factors—at a certain idea. Here we have an *absolute* conceptual tendency of an environmentally invariant character. (The idea of *self* may be an example.)

2. Propensities on the part of men to arrive at a certain idea under suitable environmental circumstances. Here we have a *conditional* conceptual tendency of an environmentally dependent kind. (E.g. ideas of up/down for us Earthlings living in a strong gravitational field.)

3. Propensities on the part of men relating not to the *acquisition* of ideas, but to their *processing*—i.e. propensities to work with ideas in a certain way and to make certain kinds of conceptual moves. (Think for example of the suggested propensities basic to the associationistic psychology.)

Our question can be put thus: Has man *a propensity to introduce certain ideas*—or certain *types* of ideas, or certain *mechanisms* for handling ideas—into the conceptual scheme he employs in structuring his *thought in language?* When the question is put in this Leibniz/Neo-Kantian framework, it is apparent that its answer resides in a search for linguistic uniformities of conceptually significant variety. The method of investigation at issue here must be considered in detail.

5 *The Methodology of Investigation*

The search for significant uniformities in the way in which concepts are embedded and put to work in the languages which provide the ' conceptual schemes ' for our thinking is by no means a simple and straightforward matter. When we enter into this complex arena, various prospects and possibilities lie before us, of which the following three are especially prominent:

Spurious Uniformities. It is easy to convince ourselves of such conceptual regularities as: *Men everywhere—regardless of linguistic environment—count by the use of numbers* or *In all linguistic settings subjects are capable of bearing predicates* or *Everyone who has a conception of a machine thinks of it as an artifact.* The sources of the compelling plausibility of such uniformities lies at the surface. They are straightforwardly *spurious* uniformities: although they have the seeming air of resulting from a careful survey of the cases, they are actually not *discovered* uniformities at all, but *imposed* ones. They in fact shed no light whatsoever upon the various language-systems with which they purport to concern themselves, but are immediate reflections of our own linguistic apparatus. *We would not call* a process ' counting ' if it did not involve the use of numbers; *we would not call* a term a ' subject ' that was not capable of bearing predicates; *we would not call* an object a ' machine ' if it were not an artifact. The regularities at issue thus reflect only the applicability-criteria of *our own* languages in general. One might as well tout it as a discovery that everyone's forks have tymes.

(Accidental) Empirical Uniformities. It is certainly possible, and indeed likely, that human languages in general share certain empirical features in common. It might be—let us assume it is—the case that all languages that have verbs vary them by person (1st, 2nd, 3rd; singular or plural). What of it? And indeed given the high degree of generality of many of the categories of modern linguistics it is

readily conceivable that certain (genuinely empirical) features capturable within the categorical scheme are shaped in commonly all known natural languages. Assuming this so in some specific instance it would be definitely premature to conclude anything from it that bears relevantly on the issue of ' innate ideas '. In the first place, the regularity might be a merely accidental or fortuitous one. (Suppose, say, that in all languages adjectival descriptors lend themselves to the formation of comparative and superlative counterparts.) In the second place, the regularity might reflect some strictly utilitarian consideration. (Suppose, e.g. that in all languages the equivalent of ' mother ' and ' father ' are relatively short.) An empirical uniformity in the applicability of certain concepts—could well be a wholly (fortuitous) matter that in no way makes manifest any underlying conceptual concentrations.

Functional Uniformity. The starting-point (but only that!) for the discovery of a functional uniformity is provided by the finding of an empirical regularity. But for reasons now to be explained, for such a uniformity to be *conceptually* significant the circumstances of the case should be such that the regularity is capable of being provided with a rationale of a particular sort, namely a *functional* rationale. The regularity must be rendered plausible (i.e. ' only to be expected ') by an exposition of the fact that—by virtue of the particular range of concepts at issue—the regularity is a *purposive* one, that is to say, one which serves the interests of effectiveness and efficiency in the relevant conceptual domain. (And of course we should study these uniformities not as isolated items, but with a view to their coherence—the extent and manner in which they hook up with one another in mutually supporting patterns.)

The considerations involved here are difficult and far from self-explanatory. The best way to expound the ideas at issue is to develop in some detail a significantly helpful illustrative analogy.

6 *The Computer Analogy*

In order to clarify the idea of a functional uniformity, let us once again draw and exploit the nowadays hackneyed analogy between the human mind and the modern electronic computers—not in the usual direction from man to machine to argue the existence of ' thinking *machines* ', but in the reverse direction to draw attention to certain features of *human* thought.

A computer has three importantly distinguishable aspects:[1]

(1) *Its 'hardware'*: the physical gadgetry constituting the computer as a particular piece of equipment.

(2) *Its 'logic'*: the mode of functioning of the computer. This set of capabilities and tendencies is not a physical part of the computer at all, but can be embodied in very different collections of hardware. It has to do with the *modus operandi* of the computer: the generic character and inevitable patterns of its functional responses under various kinds of relevant circumstances (if you like, the 'concepts' and the 'modes of thought' it brings to bear).

(3) *Its 'programme'*: the specific data inputs of its 'memory banks' and the operating instructions for its working procedures—with which the computer is specifically equipped in the course of being deployed upon a particular problem, i.e. its 'empirical experience' within a given area.

Forgetting for the present about the 'hardware', we note that the *functional* features relating to the computer's activities can be two sectors: (1) the first of which (its 'programme') is clearly a matter of acquired data and procedures, but (2) the second of which (its 'logic') is equally clearly a matter of modes of functioning which pertain to the computer by virtue of the characteristics of the 'hardware' make-up, and which remains essentially unaffected by changes of acquired materials (i.e. data and instructional inputs). It is this set of features of the computer's 'mental', that is to say *computational* make-up—features that may or may not be brought to the fore in the course of the group of problem-tasks set to a particular computer in the course of its particular history—that represent the 'innate' modes of 'conceptualization' and of 'reasoning' on the part of the computer.

What is being proposed, on our cybernetic approach to the issue, is that it is in the light of these features inherent in the *a priori* 'logic' of a computer—in contrast to its *a posteriori* 'programme'—that the conception of innate ideas be construed on the human side of the man/computer analogy. This analogy is not to be viewed as an 'existence argument' that demonstrates the actuality of 'innate ideas' in the sense at issue, but only as a 'feasibility argument' that demonstrates their *possibility*.

---

[1] For a good, nontechnical account of the nature and capabilities of modern computers see D. G. Fink, *Computers and the Human Mind* (New York, 1966).

But what is the special significance of functional uniformities in the context of our problem: (1) Is there warrant for taking functional efficiency as a mark of innateness? (2) And conversely, should the innate be expected to prove functionally efficient? (Understand *innateness* now in the strict sense of the 'logic of the machine'.) In briefest outline the line of reasoning here is to rest an affirmative response to (2) on the assumption of an evolutionary development, and then to hedge with respect to (1). For in this evolutionary light, we must regard functional efficiency of an environmentally invariant conceptual disposition not as a *proof*, but merely as a *presumptive mark* of innateness[1]: functional efficiency is not part of the meaning of innateness, but is here cast in the role of a symptomatic indicator for testing its presence. The argument to innateness is fraught with the same sort of difficulties that arise with respect to the 'black-box problem' of attempting to infer the intrinsic 'logic' of a computing machine from knowledge solely of its output.[2]

Taking a somewhat traditional view of the subject, the following objection might be raised: 'Even if there is such a thing as a "logic" (in *your* sense) of human thought, we humans shall not be in a position to discover it, because we cannot contrast it with any other possible "logic".' The computer analogy is important here for this reason, among others, that it affords at least one plausible basis for a reply, viz. that in the area of 'computer logics' a contrast-group has become available.

7 *The Project of Investigation*

The exploitation of this line of approach would call for a many-faceted investigation involving the following components:

---

[1] There is obviously no necessary connection between the functional and the innate—any more than psychological tendencies deep-rooted in man's make-up need necessarily be of positive value from the adaptive standpoint. In fact there is merit in the suggestion offered rather jokingly by one colleague of mine (Martin Tweedale) that one promising method for identifying a uniformity as innate when it is—its generality notwithstanding—from the angle of purposive considerations, highly irrational and adaptively negative.

[2] The man-machine analogy upon which we have drawn for our purposes is not rendered an illicit one in this context by the fact that a machine is a purposive artifact precisely because the quasi-finalistic atmosphere that prevails on the human side of the analogy in virtue of evolutionary considerations.

(1) There is first the *jointly conceptual and empirical* sector of the problem to delineate those kinds of ' conceptualising ' activities that have a bearing upon the area of ' innate ideas ' in the sense of relating to the ' logic ' of the human ' thinking machine '.

(2) There is secondly, the *purely empirical* (primarily psychological and brain-physiological) initial investigation of the functional uniformities that represent modes of operation of ' the *human* computer ', i.e. the characteristic features of its ' logic '. (We may think here of the sort of information that comes now and again from such diverse enterprises as structural—and especially transformational—linguistics, learning theory, and problem-solving theory.)

(3) Finally, there is the *primarily logico-analytical* enterprise—beginning when one certain element of the ' logic ' of the human computer has been determined—of determining *which of these* features can be characterised as more than merely empirical regularities: that is to say as *functional* regularities. Such regularities are certified as being presumptively better than merely accidental in not simply representing a *possible* way of accomplishing a given task, but a functionally *efficient* or *effective* way of accomplishing this task.

It might be remarked parenthetically that it would seem that the best prospects for a useful contribution on the part of philosophers are to be looked for in the context of this third, nonempirical but analytical and normative part of the problem—the conceptual analysis of functions and the specification of appropriate criteria of efficiency in their discharge. One of the main merits of our approach resides in its sharper discrimination between on the one hand the factual and empirical issues that are involved, and on the other those which are of a strictly conceptual and analytical character.

It should, moreover, be stressed that the problem in the proposed reformulation maintains an unbroken link with the historical tradition. In Plato, certain ideas are innate on ontological grounds, because they conform perfectly to a pre-existent paradigm. In Kant certain ideas are innate on *epistemological* grounds because they correspond to inevitable requisites for thought about things. In this light, the enterprise envisaged here has a markedly empirical foundation: dealing not with *ideas* as such, but with our language-embedded modes of ' working with ' ideas, we seek to isolate those functional patterns that are brought to light in a strictly empirical inquiry into the *modus operandi* of human thought.

8 *An (Imperfect) Example*

To mitigate a significant limitation of the discussion to this point—its primarily methodological and programmatic character—we turn now to the examination of a specific example for the sake of concreteness and illustrative force. The example is imperfect because its empirical basis is in some degree speculative (i.e. is less than solidly established). Nonetheless, the example is a very useful one, because—even if it fails (as it well may) to establish its particular result—it illustrates graphically the *kind* of result that can be attained on the approach we have been concerned to present.

Our starting-point here is provided by a significant empirical discovery relating to the techniques of human problem-solving—both on the side of (1) the informal problem-solving of individuals in the psychological laboratory, and on the side of (2) the attempts to create systematic means of solving whole classes of problems on electronic computers (the so-called 'problem-solving programs'). From the former we learn of the central importance of (i) the search for *similarities*—and above all for *regularities*—in the situations ('pattern recognition'), and (ii) the use of analogies to exploit such regularities by finding techniques successful in respect of 'old' problems as ways to tackle the 'new' ones. The underlying line of attack is the exploitation of structural regularities by means of analogical reasoning.[1]

The fundamental procedures involved in human problem-solving—namely comparison, similarity-noting, structural-pattern discernment, trial-and-error search, and 'means-end analysis'—are all basic to the 'mode of operation' of the human 'thinking machine'. Indeed, there is good reason to think that the number of basic information-handling processes involved in even the most creative forms of human thought is very small. On the other hand, the effectiveness, efficiency, and fruitfulness of these concept-utilising processes is evidenced by their equal prominence in deliberately optimised problem-solving programs for electronic computers. The capacity to compare and contrast and the corresponding wealth of concepts relating to the conceptual region covered by such English words as 'likeness',

---

[1] For a fuller description of the relevant issues see H. A. Simon and A. Newell, 'Information Processing in Computer and Man', *American Scientist*, 1964, 53, 281-300; and the first author's paper 'GPS: A Program that Stimulates Human Thought', in E. A. Feigenbaum and S. Feldman (eds.), *Computers and Thought* (New York, 1964).

'similarity', 'uniformity', 'pattern', 'regularity' together with their corresponding opposites[1]—represents the functional setting characteristic of the approach to the problem of 'innate ideas' that is at issue in our analysis. Again, another example is provided by the notions basic to means-end analysis, belonging to the conceptual region covered by such phraseologies as 'in order to', '... needed for—', and the like. In both of these cases, the finding of *de facto* use is backed by the warrant of demonstrable effectiveness in just the way characteristic of the 'innate ideas' of the functional variety with which we have chosen to concern ourselves. Note here, incidentally, a significant difference between the relevantly conceptual modes of performance and nonconceptual ones. One can, for example, be able to 'do' a dance without being able to conceptualise this—e.g. have vocabulary to characterise what goes on. Man, on the other hand (though not, of course, the higher mammals) standardly carries out such procedures as 'means-end analysis' and 'similarity-discernment' by overt employment of the appropriate concepts. The propensities at issue are thus ones not only to make use of certain conceptual processes, but also to arrive at the concepts by which man, as 'a rational animal', standardly carries out these processes.

Having drawn my illustrative example from the area of problem-solving, I hasten to add that this area is by no means to be thought of as the only potential sphere of application for our approach to the problem of innate ideas. The method can almost certainly be implemented in other contexts. For example, the powerful techniques of modern grammatical analysis—especially transformation grammar—represent another hopeful sphere of application. Suggestive evidence is coming to hand that certain grammatical categories represent functional conceptual uniformities of just the kind with which we are concerned, so that (crudely speaking) children do not purely and simply *learn* to work with grammatical structures, but in some degree come to discover them lying prefabricated to the hand (or rather tongue) of the human language user.[2] In the case of such functional regularities of linguistic provenience it should in principle prove possible to bring to bear the method for the analysis of 'innate ideas' with which we have been dealing.

---

[1] The list is reminiscent of some of Kant's 'regulative ideas', and our construction of *innate ideas* has a certain kinship with this Kantian conception. Specifically, the metaphysical (ontological) aspects of the innate idea tradition have been cast aside.

[2] Cf. N. Chomsky, *Aspects of the Theory of Syntax* (Cambridge, Mass., 1965), see especially chap. i, sect. 8. Because of the generality of the 'grammatical' concepts involved here, the prospect remains somewhat clouded.

9 *Conclusion*

In approaching our conclusion, it may be useful to cast a brief retrospect over the ground that has been covered. The problem of innate ideas was reformulated in terms of an inquiry into innate human propensities to arrive at and make use of certain 'ideas' or mechanisms in handling 'ideas'. These propensities are to be sought in the first analysis in those empirical uniformities in human thought processes (or—if you prefer—in the conceptual schemes inherent in natural languages). But such regularities are germane to our interests only if they prove in the final analysis to be not merely accidental or haphazard uniformities, but *functional* uniformities—i.e. features of the 'logic' of the human 'thinking machine' that reflect the attainment of demonstrably efficient and effective modes of functioning in the context of the conceptual problems that confront us. The endeavour was made to illustrate this mode of analysis with reference to the fundamentally analogical structure of human problem-solving processes.

In closing, then, it can be said that we have sought to restate the problem of innate ideas in such a way that the link with the historic forms of the issue remains unbroken, while at the same time exploiting a purely modern perspective—the man/computer analogy—both to elucidate the conceptual anatomy of the issues, and to resolve the question into a more meaningful and tractable form.[1]

University of Pittsburgh

---

[1] In writing this paper I have profited substantially from the comments and criticisms of my colleagues in the Philosophy Department of the University of Pittsburgh.

Brit. J. Phil. Sci. (1966), Volume 17, Number 3, 219–231
Printed in Great Britain

THE ACHILLES PARADOX AND TRANSFINITE NUMBERS *

C. DAVID GRUENDER

I

THE paradoxes of Zeno must surely be regarded as one of the most persistent intellectual puzzles man has invented, if we can ignore those of society itself. Sometimes they are solved, sometimes they are ignored, and sometimes they are solved anew. They do not present the same problems to everyone. My purpose in discussing a pair of them again is not to propose a wholly unique solution, for I think the general outlines of a solution are available. And because the problem involves the subtle and often evasive character of the application of abstract and mathematical concepts, I shall try particularly to disentangle these issues.

Achilles' familiar plight is traditionally put as resulting from a race with a tortoise. Fairness, of course, requires that the tortoise be given some handicap, so he starts out alone. When he reaches the agreed point, Achilles then leaves the starting line, and, as everyone knows he must, quickly passes the tortoise and wins the race. But he ought not to, for before he can get to the finish line, he must get to the point the tortoise was at when Achilles started. Yet while Achilles is reaching this point, the tortoise is reaching a point a little beyond this, which it is Achilles' next task to reach. But of course upon his arrival there, he finds the tortoise to be at a third point ahead of him, and so forth indefinitely. As Aristotle puts it:

> . . . the slowest will never be overtaken by the swiftest; for the pursuer must first reach ((the point)) whence the pursued started, so that the slower must always be some distance ahead.[1]

---

* Received 12.xi.65. A version of this paper was read to the Western Division Meeting of the American Philosophical Association in Chicago, May 1965. It has benefited greatly from discussions with Hans Herzberger, Isaac Levi, J. Sayer Minas, Raymond J. Nelson, Bayard Rankin, Leo Simons, and Richard Vesley.

[1] *Physics*, 239, b 14, trans. Gregory Vlastos, 'Zeno', in Walter Kaufmann (ed.), *Philosophic Classics*, Prentice-Hall, Englewood Cliffs, 1961, 1, 40

We may reduce our problems by considering a simpler version of the same argument: '... there is no motion, because the moving ((body)) must reach the midpoint before it gets to the end.'[1] Calling this argument the Racecourse, it needs only to be amplified to see its relation to the Achilles. It takes what is only apparently a more radical position: not that Achilles cannot beat the tortoise, but that no race could be run at all. For before one can reach the end one must get half way there; once half there one must get half the remaining distance, and so forth indefinitely. Hence no race could ever be finished; since the racecourse may be of any length, no distance can ever be covered. And since motion is the traversal of distance in time, no motion is possible.[2]

There are some important logical differences between these two paradoxes, as well as the drama of two in a race, but they share an outstanding feature; that the apparent necessity of taking an infinite number of steps seems to preclude the possibility of crossing even a short finite distance. This is the heart of the two paradoxes, and it is this problem I try to deal with. I wish, first, to show why I think some common solutions leave us still facing that problem, then, I shall try to show why the problem seems so plausible, which will be the same as to see it is only that, and last I shall discuss some auxiliary problems of the continuity of space and time.

II

The solution one proposes will then depend, in part, on what one sees as the difficulty in the paradox. And here it becomes necessary to explore the possibilities beyond what we can with any confidence ascribe to Zeno himself. One possibility is simply the difficulty that to get to the end of the course the runner is required to perform an infinite number of steps. But, one can argue, an infinite number of steps cannot be taken! Two reasons can be offered for this. One is usually put as a physical reason: there is just not enough time to complete the task; one would need an infinite period of time to take an infinite number of steps. But, as Aristotle has explained, an infinite period of time *is* available, ' for divisions and subdivisions of the given time and the given magnitude can always be made to keep pace in number and ratio without limit'.[1] That is, if you must cover a given distance in a given time, you can still whisk yourself to your destination by taking half the time to get half way, one quarter of the remaining time to cover the third quarter of the distance, and so forth, until you are there.[2] So if you *must* take an infinite number of steps to go one hundred meters, two minutes will still provide in this manner the infinite time you need. Just go 50 meters in the first minute, 25 in the first half of the second, and in two minutes you will be there.

---

[1] *Physics*, 239, b 11, trans. Vlastos, p. 35

[2] One can reconstruct Zeno's argument in other ways. For a discussion of the historically likeliest possibilities, see Vlastos, pp. 35-40. Another important version has the runner unable to start the race because before he can even get half way to the finish he must get one-quarter of the way, and so forth. But I think Vlastos' arguments for the reconstruction of it I used above are sound. And I have followed Vlastos in calling the argument the Racecourse instead of the Dichotomy.

But this reply, conclusive as it seems, may serve to force one to restate the difficulty as a logical rather than a physical one: the reason there is not enough time is that the task of taking an infinite number of steps is one, that, by definition, cannot be completed; for however many you have taken, there are always a number of them still in front of you. *There is no last step*, and without taking one your task cannot be completed.[3] If this is the spot we are in, that there is not enough time is but irony; for the task of finishing the unfinishable cannot easily be carried out.

In reply to this it has been argued that by assuming that if any number of tasks have been performed, that some task has been performed last begs the question.[4] Rather one can conceive of an infinite series of acts having been performed, so long as one means by the completion of this task not that some act has been performed last, but only that each of them as been done.[5] As an aid to the imagination here, let us introduce an infinity ' machine ' that is hindered by no physical problems. Several models have been designed. One, for example, 'works' with marbles: part of it moves a marble from a left-hand tray to a right, then tests; while it is resting the other part moves that same marble back. We now set the machine to repeat this act an infinite number of times (once each instant) by, say, pushing a button marked: 'infinity'. Two minutes should, of course, be plenty of time for it. When the machine has finished, in which tray is the marble?[1] Or take a 'machine' already in partial existence: a simple night light with a switch which pushes in; push it in and the light turns on; push it in again and it turns off, and so forth. Now push it an infinite number of times. When you are through (two minutes later) is the light on or off?[2]

---

[1] *Physics*, 239, b 20. This trans. Wicksteed and Cornford, Harvard, Cambridge, (Loeb Classical Library), 1934

[2] Many writers since have made the same point. Among them, see Bertrand Russell, *Our Knowledge of the External World*, rev. edn., Allen & Unwin, London, 1926 (1st edn. 1914), pp. 177-178; H. Weyl, *Philosophy of Mathematics and Natural Science*, Princeton, 1949, p. 42; A. N. Whitehead, *Process and Reality*, Macmillan, New York, 1929, p. 107.

[3] Max Black, ' Achilles and the Tortoise ', *Analysis*, 1951, **11**, 91-101, republished in his *Problems of Analysis*, Cornell, Ithaca, 1954, as chap. vi. See section 10 of this paper.

[4] Richard Taylor, ' Mr. Black on Temporal Paradoxes ', *Analysis*, 1951, **12**, 38-44

[5] J. Watling, ' The Sum of an Infinite Series ', *Analysis*, 1952, **13**, 39-46

But ingenious as these machines are, the reply may seem defective because the earlier analysis showed that it would never make sense to ask where the marble is or whether the light is on, for these questions imply that there was a last act. Such an implication is not only incompatible with the notion that the action was performed an infinite number of times, but is unnecessary to the completion of the task, which requires only that all of its parts be performed, not that one of them be performed last. That is, a task may be regarded as completed when each and every one of its parts has been performed—when *all* of them have been performed. If one is asked to count all the odd numbers, he is not thereby told to count one of them last, much less how to count them. And however he does it, I think we must count his task finished when they are—all of them—counted. We may not conclude anything at all about whether or not he counted one or another of them last. That is, if a task consists of the set of parts $p_i, p_{i+1}, \ldots p_{i+n}$, it is completed when each of the parts p_n has been performed. And that they have all been performed does not itself imply that one of them was performed last.

And yet: Where is the marble? Is the light on or off? It is all very well to say that this is a meaningless question, but in the physical world the marble has to be somewhere and it has to have gotten there somehow,[3] just as the light has to be on or off, the switch closed or open. The question that is asked is not which act was last, but: given the completion of the task in the weakened sense of 'completion' proposed, what then is the state of affairs? This last is an appropriate question. Failure to answer it may be taken as failure to provide an adequately clear meaning to the notion of completing an infinite series of discrete acts.[1] Or it may be taken as showing that the notion is simply incompatible with our concepts of the physical world.[2]

---

[1] Black, section 17

[2] J. F. Thomson, 'Tasks and Super-Tasks', *Analysis*, 1954, **15**, 1-13

[3] Max Black, 'Is Achilles Still Running?' in his *Problems of Analysis*, chap. vii, p. 111

The first is a charge of unclarity; the second, of empirical falsity. It seems to me that both are correct, but for reasons that are very much stronger. The notion of the completion of a task consisting of a *finite* series of discrete parts is certainly clear. It is of course a physical task, and each of its parts is successively performed in time. Since each part is discrete and follows serially upon a previous part and they have all of them been performed, it follows that some one part *has* been performed last, both in this order and in time. If we now say that the task in question has an *infinite* number of parts, we have put together two incompatible notions. For each member of the discrete successive series or parts to have been performed, requires that some part be performed last in the serial order. Yet that there be an infinite number of such parts requires that there be no last part in the order. This incompatibility is not at all removed by weakening the meaning of 'completion' to imply merely that each act be done, for their performance in a discrete serial order still requires that if each of them is done, some one of them is last in that order. And it is in just such an order that the marvellous marble-mover, the prodigious night light puncher, and the runner pacing along a racecourse take their steps. Thus the notion of completing an infinite series of discrete acts is unclear because it contains incompatible elements. And because it does, it certainly will not fit our concepts of the physical world.

Let us take another direction and apply the principles of the calculus to the problem; they have been used since their discovery for the successful analysis of motion. One thinks of the distance to be covered and the time which this is to take as divided up into very small elements —'as small as you please'. Let Δs be the distance covered in the time interval from some time t_0 to some later time $t_0+\Delta t$. Then the limit of the ratio $\frac{\Delta s}{\Delta t}$ as Δt approaches 0 is the instantaneous velocity at time t_0. This limit, it is attractively put, may be 'approached as closely as you please'.

So you see one can get from one end of the racecourse to the other in a definite time at a definite velocity. Had Zeno only known some elementary calculus he would have had the solution to his problem! This kind of statement is not made as often as it was formerly.[1] Its apparent plausibility may lie in the wording 'approached as closely as you please', for this phrase does *not* say that the limit is *reached*. And if the limit is not actually reached, we have not got a velocity by means of which we may get to the end of the racecourse, which is to concede Zeno's point. This point remains unaltered by the suggestion that we have done something every bit as good by getting 'as close to it as you please', for what is required is simply that you reach the velocity, and sidling towards it at your pleasure is not reaching it.

---

[1] Thomson, p. 6

[2] Black, ibid. p. 112

It is important to remember here, that the limit here referred to is the instantaneous velocity, not the end of the course. But it ought not to have persuaded Zeno of his mistake because it offers an analysis of the runner's motion in precisely the terms the paradox is calling into question. The modernity of this way out of the difficulty is sometimes overstressed too, for the method of exhaustion to a limit, although it may not have been developed until after Zeno, was nevertheless a part of Greek mathematics before Archimedes' masterful use of it.

The situation is analogous if we look at it from the point of view of infinite series. The process of successively halving the portions of the racecourse to be traversed generates the infinite series

$$1/2,\ 3/4,\ 7/8,\ 15/16\ \ldots$$

This series has no last member, but it can be shown to converge toward a limit, and thus we may define what is called the 'sum' of the series:

$$1/2 \text{ plus } 1/4 \text{ plus } 1/8 \text{ plus } 1/16\ \ldots$$

as the limit which is approached as we take the 'partial sums' of greater numbers of terms. But the difficulty is just as it was before. This 'sum' of the series is not really a kind of sum: the specified additions cannot be completed, so although the limit can be approached, it cannot be reached.

And at that it must be pointed out that by setting up the series by a process of successive halving, we are giving the opposition the best possible run for its money. For that series does at least converge toward a limit which is the finish line. But one could, with as much justice, construct an indefinitely large number of other series which either do not converge at all, or converge toward a limit short of the goal.

---

[1] Cf. S. Sambursky, *The Physical World of the Greeks*, trans. Merton Dagut, Collier, New York, 1962 (trans. 1st published 1956), pp. 175–176

For example, one could say that before you can get to the end of the course you have to get one-third of the way, then one-third of the second-third, and so forth. This generates:

$$1/3+1/9+1/27\ldots$$

whose ' sum ' is not 1, the finish line, but 2/3, which is one-third short of it.[1] But this is only to say that no point on the track can be reached.

However, if you think about the problem as the physical task of a human runner, another solution may appear: the actual success of the runner shows that the abstract mathematics of the calculus or of infinite series does not really fit physical space and physical time.[2] ' The infinite belongs not to the race, but to its form of representation. . . .'[3] It is just physically impossible for a human runner to take actual steps which are successive halves of the remaining distance. The runner just takes a finite number of paces of more or less equal size and reaches the end of the course in a finite time. There is, of course, nothing in the *logic* of the situation (to which the *physical possibilities* are irrelevant) that keeps us from talking in terms of ever-diminishing steps. Nor can there be any harm in this so long as we remember that we are talking about two kinds of steps: physically separate and logically separate. One can get to the finish line by taking a finite number of physically separate steps. And when one has done this, he has covered ' all ' of the infinitely many logically separate steps of the successive halving—and more!—for one has gotten beyond them to the finish line.[4]

So far nothing has been said that one can directly quarrel with. The distinction between logical and physical acts is an important one, and there can be no doubt that successive halving is physically impossible if carried very far. One can devise gadgets that could carry it farther than people, or well-trained ants,[5] but that, I think, misses the point. At the same time, there remains something very puzzling here even after this. For the logic that seems to permit the physically impossible successive halving is not completely abstract. It is not the logic of *n*-dimensional space, nor that populous realm of logical possibility conjured up by non-self-contradictory sentences where dwell the fabled unicorn, the gorgon, and the manticore; those marvellous men who do wear their heads beneath their shoulders; and other wonders as yet undreamt. Quite the contrary; the logic that permits the impossible is a part of our conceptualisation of the space and time in which races are run. And I think this is not avoided by saying that these concepts are mere conventions, and the complexities belong to them, not to physical space and time. If they permit the impossible, there is something wrong with out concepts.

---

[1] David S. Shwayder, ' Achilles Unbound ', *Journal of Philosophy*, 1955, 52, 451

[2] G. J. Whitrow, *The Natural Philosophy of Time*, Nelson, Edinburgh, 1961, p. 148; Black, ibid, p. 126

[3] Shwayder, p. 453. He goes on to say that we can avoid the problem simply by not using the word ' infinite '; all that we mean, really, is that ' we have adopted a certain kind of language . . . such that . . . it is nonsense to ask for the last step '. But since the problem hinges on that very impossibility, this seems to me to be verbal sidestepping.

[4] Vlastos, pp. 39–40

[5] J. A. Robinson calculates that a computer could not go much beyond thirty-five steps, in his ' Zeno's Paradoxes and Computing Machines ' read at the University of Houston, 1961 (partially printed in *Zeno and the Texas Limited*, ed. Robert Rein'l, Southwestern Philosophical Society, 1962).

III

Let us give them closer attention. In what has gone before I have used the word 'infinite' without qualification, but we should now use Georg Cantor's analyses of the notion of infinity. The discrete steps corresponding to the infinite series:

$$1/2,\ 3/4,\ 7/8,\ 15/16\ldots$$

form a transfinite set of the lowest order (aleph-null). It is bounded on the lower end, but has no last member and thus no upper limit that is a member of the set, and is 'denumerable' since we could map it against the counting numbers. The racecourse itself, however, forms a spatial continuum. It is bounded on *both* sides, has infinitely many points between the two sides; so many, in fact, that there are infinitely many between *any two* points—hence its 'density'. This infinite collection of points is 'non-denumerable'. It is also 'connected' in that one cannot choose two sub-sets of points each of which exclusively contains its own end-points. If Cantor's continuum hypothesis is correct, it has a higher-order transfinite cardinal number (aleph-one); but this continuum has at any rate some cardinal number *c*. Now what is not so clear before this kind of analysis of infinite numbers is that our concept of a continuous manifold defined in terms of a second order infinity has one quality of overriding importance for our problem: it permits the construction of discrete infinite series of the lowest order within it. For if there are an infinite number of points between any two points, however close, there is 'room' and to spare for ' all ' the end-points of the lowest order infinite set generated by successive halving to fit between them. I say ' all ', of course, because since the series has no last member, one cannot speak of ' all of its members ' in the sense in which we can speak of all the members of a finite set, which *can* be arranged so that one member is last; ' all ' here can mean only that no member of the infinite set is left out.

Because the very notion of a continuous manifold of any sort bounded on both ends by members of itself permits this construction within it of *aleph-null* infinities, which are discrete and unbounded on one ' end ' by their own members, our original problem arises. For it is this I think that, unnoticed, generates the plausibility of agreeing with Zeno that before a runner can get to the end he must first get to the mid-point, then three-quarters of the way, thus launching him on his hopeless attempt to traverse a denumerably infinite sequence. It is hopeless both because his task has no last member and because his destination—the finish line—lies outside this *aleph-null* infinite sequence.

The paradox is resolved not only by seeing the two different orders of infinity that are here, but that the runner's task is to traverse the higher order infinity, the doubly-bounded *Cantor-c* continuity of the racecourse. He can, of course, do this physically in a finite series of temporally continuous or discontinuous steps, the continuity of the course permitting him to break it up into portions or intervals as he may. If we give him a bicycle he can cover it in a completely continuous manner: legato, not staccato, as has been urged.[1] Although whether on wheels or on foot, the smoothness of his performance or the lack of it, must not be confused with the continuity he is to traverse.

Yet in whatever manner he covers it, within the continuous manifold that he traverses from one end to the other, our concept permits us to construct denumerably infinite sequences which contain no upper limit. Since he traverses the continuous racecourse, it appears that we may say he also ' traverses ' each of these *aleph-null* sequences, although ' traverses ' is here used in an extended and subsidiary sense, for it can usually only be applied to the crossing of a self-bounded interval. We may think of them as ' traversed ' because the continuous interval is traversed, and we may construct them logically within it. But although they can be constructed within the continuous interval, they are not its parts, for the interval is not constructed of them: their sum does not exhaust the continuous interval. Hence in carrying out the task of traversing a racecourse, a runner incidentally 'traverses' denumerably infinite sequences. But no attempt to traverse (in the usual sense) any such sequences is logically possible; neither could it constitute the traversal of a continuous interval. It could, therefore, hardly be said to constitute a necessary condition for that latter task.

---

[1] Adolf Grünbaum, ' Modern Science and the Refutations of the Paradoxes of Zeno ', *Scientific Monthly*, 1955, 81, 238

IV

If this diagnosis of the paradox is correct, its heart lies in misunderstandings about logical relationships between concepts of infinity. Much of the discussion about it, however, has also dealt with the relationship of these conceptual schemes to reality. This latter is of interest not only for this particular case, I think, but also for what it illustrates about problems of this kind generally, which are tricky at best.

One view that has been widely held is typified by Bergson's contention that the flow of time can be directly experienced by any reflective person, and that when it is, it becomes immediately apparent that its smooth continuity is primitive and cannot be further analysed. For to analyse it would require one to break up its continuity into discrete elements; yet to do so is to falsify its intuitively evident continuity. Thus does Reason betray Nature by its blind urge to geometrise.[1]

There are many issues here on which I shall not now comment. I think Bergson, James, and others are right in insisting on the continuity of *experienced* time.[2] I would maintain that it does not follow from this that *physical* time has the same character, but whether this be granted or not, I think it obvious that physical time is continuous if we confine the discussion (as in all the foregoing) to the macroscopic world of ordinary and everyday events and their analyses on this level. For an interval of time is subject to measurement at any point and an interval of any ordinary finite length is possible. A runner can reach the finish line at any time that his physical abilities permit, and a timer can squeeze a stop-watch whenever he gets there—he is unhindered by the discontinuous jerks by which the hands of his mechanical watch move. And racecourses, too, are continuous—on the macroscopic level. However the level at which one wishes to say that some manifold is continuous or is instead a series of discrete elements with gaps between them, depends on the purpose of the analysis. For the purpose of a runner, the gaps between the cinders in the track, although discoverable by the naked eye, do not affect the continuity of the racecourse. But for that same purpose the continuity of the track would be effectively destroyed were an earthquake to open a two meter crack in it.

---

[1] Henri Bergson, *Creative Evolution*, chap. I

[2] Cf. Grünbaum, p. 237; Whitehead, p. 95

It is the possibility of being at any point or stopping at any point in a spatial or temporal interval that I take to be the essential feature of its continuity, and hence of the applicability to it of a cardinal *c* infinity. If this analysis is adequate, then there is no doubt that continuity can be analysed, in a strong sense of ' can '. (For not all analyses must be in terms of discrete elements, whatever might be suggested to us by the use of this term in simple chemistry.)

Because of the same feature of macroscopic physical actions occupying any physically possible point of space or time, no further demonstration of the continuity of ordinary spatial and temporal intervals seems required. But suggestions to this effect have been offered. Mr. Grünbaum, for example, although taking the view that Achilles has only to traverse a *Cantor-c* continuity of space and time, agrees with those who regard the experience of time as discrete, like ' beads on a string '. He therefore proposes to substitute a physical criterion for the passage of time that will make plain its continuity. His criterion, a modification of proposals by Boltzmann and later by Reichenbach to ensure the flow of time in one direction, is: that, of two entropy states of a closed non-equilibrium system, the state of higher entropy be called the later one. Although some important qualifications are necessary here, Mr. Grünbaum goes on to point out that the values of such entropies are given by real numbers and hence can ' characterise a set of physical states that form a linear Cantorean continuum and are thereby ordered with respect to the relation " later than " '.[1] Although it is clear that a *Cantor-c* continuum can be provided by this device, the difficulties of determining whether the energy distributions called for can be satisfied are enormous.[2] But whether these difficulties could ever be resolved or not, the operations of determining energy distributions of systems of physical states is not generally what we can do or mean to be able to do when we talk about the continuity of temporal duration. I think all we mean by its continuity is that we can be anywhere in it. And this applies to spatial continuity as well, which is not explicated by thermodynamic considerations.

[1] p. 238

[2] See Whitrow, pp. 284-288

Another element in the discussion that has frequently appeared has been the evident physical impossibility of progressively reducing the size of one's steps as called for in the formula of successively halving the remaining distance. The problem has sometimes been located here, whether one thinks it *logically* possible to traverse a denumerably infinite series or not. Although it is logically impossible to traverse such a series, since it has no last term, one should nevertheless note that the practically evident physical impossibility of carrying the series very far—its 'merely medical impossibility', as Russell once put it—brings one to face the question of whether this is the result of an inadequate technology, or whether there is a minimum physical distance or time interval. A number of writers have speculated about this possibility, and Weyl has even proposed names for them: 'hodon' for the spatial, 'chronon' for the temporal interval, and dimensions based on an extrapolation of the limits of action in quantum mechanics.

If there is a minimum length and minimum physical time interval, that this were the case would be an empirical matter of fact. The existence of such lower limits would certainly stop a runner from progressively halving his steps at some definite point, and it is safe to say that this would be before he carried out the logically impossible task of reaching the last member of the series. Then he would have both a logical and a physical excuse for his failure. But of course he would have an excuse based on human anatomy far before he could hope to hop in hodons (if such there be). In fact minimum times and distances of these orders of magnitude would not affect our runner at all; for his purposes the track remains continuous. At micro-microscopic levels, however, the effect would be to deny an essential condition of continuity, and with it the possibility of applying either the axioms of continuity or the construction of denumerably infinite subseries. Without either of these, of course, the problem cannot arise.

The possibility so far is not more than speculative, but if realized would reduce the number of apparent anomalies now troubling physical theories. Without commenting on these, it would be less than fair not to point out that what is crucial here is an important matter glossed over in the discussion, above, on the conditions of continuity in ordinary macroscopic spatial and temporal intervals. Although insisting on universal localisability, the qualification 'physically possible' was always insinuated, and I must stress that these are weasel words. They are plausible only in a situation in which we have no significant doubts about the limits on physical action imposed by the relevant physical theories, and all apparent discontinuities can be explained by these theories. We can then understand, for example, the discontinuity the earthquake produced in our racecourse as a physical matter; we are not tempted to wonder about discontinuities in space or time. But that is not at all the case with quantum mechanics, so that it is not clear whether one ought sometimes to talk about physical impossibilities, or about spatial or temporal discontinuities, or how you can tell when you have one or the other.

Philosophy Program of Western Reserve University
and Case Institute of Technology,
Quail Building,
Case Institute of Technology,
University Circle,
Cleveland, Ohio 44106

DISCUSSIONS

Can a Machine be Conscious?
Discussion of Dennis Thompson

In a recent paper,[1] Dennis Thompson makes the following claim: ' If we refuse to regard as conscious a robot which can do everything a human can, we (1) strengthen a certain kind of solipsism, and (2) commit ourselves more fully to Epiphenomenalism than we would if we rejected the Machine Theory.' (The Machine Theory is simply the view that machines *can* be conscious.) In support of the first point he argues,

> By saying that a robot can do everything a human can do, we imply . . . that the robot has all the characteristics which justify our calling humans conscious. Thus, if we deny that robots are conscious, as the Anti-Machine Theorists would have us do, we have less reason to say that other humans are conscious. The denial of the Machine Theory thus undermines our justification for the belief in the existence of other minds (p. 42).

In support of the second point he argues,

> By rejecting the Machine Theory we are denying that robots have minds, even though they can do everything a human can do. But if a machine which can do everything a human can do does not have a mind, then there is no reason to suppose that the minds of other people have causal efficacy in the physical world, for it follows on this view that we can account fully for other people's behavior without reference to a mind. Moreover, if the minds of other people have no causal efficacy, then I have little reason to believe my own does (p. 42).

I wish to examine these two arguments in order to point out what there is about them that is, it seems to me, fundamentally mistaken.

Concerning the first argument it can be said that here Thompson makes use of a most questionable assumption regarding the criteria we invoke when we attribute consciousness to a being other than ourselves. Early in his paper Thompson suggests that ' the only way of deciding whether another person is conscious is by observing his behaviour and his bodily state ' (p. 35), and thinks that it follows from this that ' If performance is sufficient justification for calling humans conscious, then it is surely sufficient justification for calling machines conscious—even if consciousness is regarded as a property of some sort of " inner " life ' (p. 35). Now it if were actually the case that ' performance ' were the sole criterion of consciousness, then Thompson's conclusion would indeed follow. But this, it seems to me, is clearly not the case. Where human beings are concerned, there is, besides similarity of behaviour, another kind of similarity that is totally absent in the case of machines, and it is *this*—and, I would argue, *primarily* this, not performance—which leads us to attribute to other people the kind of consciousness which we know ourselves to possess. The similarity in question is, to put the matter simply, the biological or physiological similarity between ourselves and other humans. Since we can discern not only that they behave in a manner similar to the way we behave, but also—and more importantly—that they have the same kind of physiological constitution as we ourselves possess, we simply have no *reason* to doubt that other people are, like us, conscious beings. If we did not know that those other beings whom we call human do in fact have a physiological constitution more or less identical with our own—the same essential organs, tissues, glands, nervous system, etc.—we might well wonder whether we were justified in attributing consciousness to them, just as we wonder now about machines. But since we do know this to be the case, the question does not arise; nor is there any reason why it *should* arise.

---

[1] This Journal, 16, 5

With respect to machines, however, the situation is totally different. Here we have a kind of being which is *given* as having a quite different physical constitution: instead of a nervous system there are wires, tubes, and rheostats; instead of a brain there is a computer; instead of a heart, a power supply, and so on. Consequently there is, in the case of machines, ample reason to question whether—performance notwithstanding—a machine is the kind of thing which *could* be conscious. There are simply too many important differences—purely physical differences—between ourselves and machines not to raise doubts on this matter. It is conceivable, of course, that a kind of ' synthetic person ' could be constructed, replete with artificial heart, liver, blood vessels, etc., but even so there would be no genuine similarity between the physical constitution of such a being and a human—unless, perhaps, artificial parts could be manufactured that were *precisely* like human organs (tissues, glands, etc.), and these could be integrated in the organic way they are integrated in a human being. If this should turn out to be possible, the sense in which such a being could be considered a ' machine ' is not clear; but at any rate it is doubtful whether many of those who are now opposed to the Machine Theory would balk at attributing consciousness to such a being. And if anyone would do so this would almost certainly be due to his inability to decide whether such a ' borderline ' person is really *enough* like human beings in all pertinent respects to remove doubts about his being the kind of being to whom a mind can be attributed. A machine must, by its very nature, necessarily lack some human characteristic (even if it is simply the characteristic of not being manufactured). This being the case, a question can always arise as to whether a machine is sufficiently humanoid to justify regarding it as conscious. Since such a question cannot arise in the case of other persons, it follows that a machine, no matter how technologically perfect, cannot have ' all the characteristics which justify our calling humans conscious '.

Thompson's second argument seems to me to be perhaps even less convincing than the first. Here he claims not just (1) that rejection of the Machine Theory has important implications for our view of other persons, but also (2) that it has considerable significance for our view of *ourselves*. Even the first claim, however, seems ill-founded. For from the fact that a mindless *machine* might perform in a manner indistinguishable from the ways in which humans behave, it certainly does not follow that ' we can account fully for other people's behaviour without reference to a mind '. As we have seen, on the basis of physical differences alone, a machine is a different *kind* of thing from a human being; consequently, there is no reason to suppose that what might be true of a *machine* is also true of a *human*. It could be the case that though their behaviour is identical, a machine could behave in the way it does without a mind, whereas a human could not. Nor does it follow that if a mindless robot could do everything a human can do there would be 'no reason to suppose that the minds of other people have causal efficacy in the physical world'. For not only can I not conclude that what is true of a robot must be true of a human, I have quite *good* reasons for supposing that other people's minds have causal efficacy in the physical world. For—and this brings us to Thompson's second claim—I have what seems to me virtually conclusive evidence that *my* mind has causal efficacy in the physical world, and other people are, so far as I can determine, exactly like myself. Furthermore, even if it were the case that the existence of robots behaving exactly as humans behave required me to admit that the minds of *other people* have no causal efficacy in the physical world, it would by no means follow that I have 'little reason' to believe that my own mind does. Whatever may be the case about machines or other minds, the first-hand evidence I possess of the operations of my mind assures me that it does have this type of causal efficacy.

In general, the situation we are in with respect to both our own minds and those of other people is quite other than Thompson suggests. Because we have *no reason whatever* to question that other people have exactly the same physical constitution as ourselves, and because we see them behaving in a manner similar to the way we behave, we attribute to them the kind of consciousness we ourselves possess. It is the two facts together, not just the latter, that leads us to do so. The case of machines, however, is quite different. So different from us are they—and this includes the humanoid type of robot Thompson has in mind as well as more conventional types of machines—that it is questionable whether we even know what it would mean to attribute consciousness to such beings. But whether we would or not, the fact is that in the case of machines, there is very good reason indeed to doubt whether they are beings like ourselves, regardless of how they perform. Since, then, we have genuine reasons for questioning whether machines can be conscious, reasons absent in the case of other people, I conclude that we cannot extend our conclusions concerning the former to the latter or, a fortiori, to ourselves.

ROBERT J. CLACK

ORIGIN AND CONCEPT OF RELATIVITY:
REPLY TO PROFESSOR POPPER

PROFESSOR POPPER[1] does not explain how in spite of 'scrupulous attempts at fairness' and 'obvious scholarly care' (high praise, thanks to Sir Karl), I arrive—apparently vagariously—at a mistaken assessment of Einstein's (E's) work. Could it be that Popper is not able to detach himself from the past?

My point is rather simple. The theory in question principally meant two things: impossibility of detecting absolute motion, limiting velocity equal to that of light. Poincaré (P) and Lorentz (L) said these things before E. The mathematical instrument of the theory, Lorentz transformation equations (L.T.E.), was fashioned by L (supposedly) to meet these two requirements.[2] E knew possibly all and at any rate most of these earlier ideas and methods before he wrote his own paper (submitted personally on 30th June, 1905; published on 30th Sept., 1905).

---

[1] This *Journal*, 1966, 16, 332

[2] *The Principle of Relativity*, Dover, pp. 11-14

Popper says that P's paper of 1905 (published on 5th June) contained only an inspiring programme sketched for a relativity theory (faint praise). It contained the following: reference to negative results of experiments undertaken to measure the absolute motion of the earth; it is a general law of nature that absolute motion cannot be demonstrated; L.T.E.; L.T.E. form a group; invariance of Maxwell's full equations under L.T.E.; L.T.E. mean that absolute motion cannot be detected; 'relativistic' velocity-addition formula,[1] gravitational waves are propagated with the velocity of light. Earlier, P had enunciated the principle and the second postulate. Oh! Will a Daniel come to judgment?

There remains the disreputable but useful-when-in-difficulty F-L contraction of lengths. In 1904, L[2] rather took the view that the contraction (ostensibly relevant) failed to explain the experiment of Rayleigh and Brace[3] and set himself the task that it is necessary to show by certain fundamental assumptions that 'many electromagnetic actions were entirely independent of the motion of the system'. He then postulated L.T.E., recognising the limiting velocity of light. As to P,[4] he objected to the contraction hypothesis sometimes.

It is said that the length-contraction is reciprocal and merely a question of 'perspective', according to E. (There is no explicit statement to this effect in E's paper.) Let Popper consider the passage in E's paper about time-dilation instead, when this perspective is lost and there ensues the 'peculiar consequence', as E[5] put it, that a moving clock goes (not 'appears'; there can be no difference between 'appearing' and 'being' in the case of facing clocks) slower than a 'stationary' clock. According to E, then, this was a velocity (not acceleration) effect.

Moreover, the important matter is what a theory explains and predicts. And what was the difference between P, L and E in this respect? None of the three ever cited an experiment which supported his notion, whatever it was, but not that of the other.

Lastly about my comments on the following errors in E's original paper[6] which Popper disposes of in a small footnote, saying that I misunderstand the passage, etc.

Formula for dependence of mass on velocity: There is no misprint. The corresponding preceding equation is also incorrect $\left[m\beta^2\frac{d^2y}{dt^2} \text{ or } m\left(1-\frac{v^2}{c^2}\right)^{-1}\frac{d^2y}{dt^2}=\text{Force}\right]$. In fact the whole of § 10 of E's paper is difficult to comprehend because a *moving* mass within each frame is *initially* taken as *constant*, the equation of motion used being: $\text{Force} = m\,\frac{d^2x}{dt^2}$, etc.

---

[1] *Oeuvres de Henri Poincaré*, Gauthier-Villars, 1954, p. 490. I failed to acknowledge Poincaré's priority in this respect, in my paper.

[2] *Ibid.* pp. 11-13

[3] Changes in lengths are not to be regarded as strains, on account of resistance. The lengths—and clocks—have a natural measure corresponding to motion relative to the universe. Now, some people find it easy to imagine accelerations, relative to the universe, retarding all types of clocks equally, but not uniform translatory motion!

[4] See remark of Lorentz in *The Principle of Relativity*, Dover, p. 13

[5] *Ibid*, p. 49

[6] *The Principle of Relativity*, Dover, p. 61-63, 57, 44

Equation of expanding wave-front of light: E gave the equation of the wave front of the light-complex observed in the moving system, $k\ (\xi,\ \eta,\ \zeta,\ \tau)$, as:

$$(\beta\xi - l\beta\xi v/c)^2 + (\eta - m\beta\xi v/c)^2 + (\zeta - n\beta\xi v/c)^2 = R^2,$$

where l, m and n are the direction cosines of the wave-normals. Mark the first β in this equation of an ellipsoid. H. B. Levinson[1] has also commented on this mistake recently.

Derivation of L.T.E.: Of course, E used velocities of light propagation *observed* in the 'moving' system as $(c-v)$, $(c+v)$ and $\sqrt{(c^2-v^2)}$, contradicting the second postulate. Popper cannot simply say that I misunderstand E, without a word of enlightenment on a correct understanding.

The situation, as I submitted in reply[2] to Dingle, was that all the three pioneers were—naturally—groping their way. Groping continues, more's the pity.

G. H. Keswani

Room No. 28 A, Annexe I
Baroda House, Curzon Road
New Delhi, India

---

[1] This *Journal*, 1965, 16, 246. I do not, however, agree with all that has been said here.

[2] This *Journal*, 1966, 17, 149

REVIEWS

THE STRUCTURE OF SCIENCE

(1) SCIENCE, says Professor Nagel,[1] 'takes its ultimate point of departure from problems suggested by observing things and events encountered in actual experience' (78); 'it aims to understand these observable things by discovering some systematic order in them' that is, by explaining them (78). It is this 'desire for explanations which are at once systematic and controlled by factual evidence that generates science' (4; 15). Understanding the structure of science, therefore, means understanding the structure of scientific explanations (15).

The view of explanation set forth by Professor Nagel does not differ in any essential respect from the hypothetico-deductive account. However a considerable amount of detail is added. In what follows I shall deal with the distinction drawn between experimental laws and theories (chap. v); the cognitive status of theories (vi); the explanation of established theories (laws) by other theories (chap. xi); and with Nagel's general views concerning the structure of science.

(2) Theories are distinguished from experimental laws because they 'employ terms like "molecule" which ostensibly designate nothing observable' (80). The distinction does not mean that laws are more highly confirmed than theories, or that theories are 'entirely speculative' (80). Nor is it assumed that experimental laws express 'relations between data . . . apprehended directly or non inferentially through various sense organs' (81; 122). The vagueness of the term 'observable' (83; 90); the difficulty of proposing a 'precise criterion' distinguishing theories and experimental laws are duly acknowledged. Yet the inference that 'the distinction is spurious because it is vague' is rejected by asserting 'several well marked features' differentiating the former from the latter (83). These 'several features' essentially boil down to one: each descriptive term in an experimental law 'is associated with at least one overt procedure for predicating [it] to some observationally identifiable trait when certain specified circumstances are realised' (83). This feature connects with others, viz. (i) the meaning of the descriptive terms of experimental laws, the so called observational terms, is partly 'fixed' by the procedure; hence, (ii) 'an experimental law, unlike a theoretical statement, invariably possesses a determinate *empirical content*' (83); (iii) experimental laws but not theories can be the result of inductive generalisations (85); (iv) to the extent to which they are connected with operational procedures the content of experimental laws *is*

---

[1] *The Structure of Science*, by Ernest Nagel, Routledge & Kegan Paul, London, 1961, pp. xiii + 618, 60s.

independent of any theory that might be used for their explanation. It also *must* be independent, 'the law must be intelligible (and must be capable of being established) without reference to the meanings associated with it because of its being explained by (some) theory. Indeed, were this not the case for the laws which a given theory purportedly explains, there would be nothing for the theory to explain' (87). Hence, 'an experimental law has . . . a life of its own, not contingent on the continued life of any particular theory that may explain the law' (87).

Theories are analysed into (1) an abstract calculus whose postulates 'assert nothing since they are statement forms rather than statements' (91) but 'implicitly define the basic notions of the system' (90; 160); (2) correspondence rules, relating theoretical notions to 'observational procedures' (94) or 'experimental concepts' (95). Such rules make it possible to use the theory 'as an instrument of explanation and prediction' (93; 106), and to investigate 'its material truth or falsity' (222). Without correspondence rules a theory is not even a statement (141) as its descriptive terms (or rather those for which no rules of correspondence are given) have the status of variables (132; 95). Finally, (3) a theory contains a model. A model is defined relative to the abstract calculus and consists in a specification of its postulates that turns them into true statements and preserves the logical relations between them (96, n. 4). A model in this sense need not provide correspondence rules (95, 97) although it can suggest 'at what point [such] rules may be introduced' (113). Nor does a model turn the postulates into statements and the theoretical terms into constants (132). For while it might contain constants only, it is not the only possible model (97). This also explains why even overwhelming evidence for a theory 'may not be assessed as sufficient for asserting the physical existence of various elements in the substantive model in terms of which the theory is formulated' (117). The calculus, the correspondence rules, the models must 'not . . . be construed as separate items, introduced in succession at various stages in the actual construction of theories, but simply as features that can be isolated for purposes of analysis' (106 f).

(3) Now it appears that these features can also be isolated in the case of experimental laws and, more generally, in the case of any set of observation statements that possesses suitable closure properties. *Observational terms* are 'associated with overt procedures for predicating them of observationally identifiable traits' or with *ostensive rules* as we shall call them.[1] *Theoretical terms* are related by correspondence rules to observation terms which in turn are related to observable traits by ostensive rules. Separating the theoretical terms from the correspondence rules leaves an uninterpreted system which can be expressed in the form of a calculus C_T with postulates P_T (91). Separating the observation terms from the ostensive rules leaves

---

[1] I have taken this very apt term from Professor Körner (*Conceptual Thinking*, p. 7).

again an uninterpreted system which can also be expressed in the form of a calculus C_O with postulates P_O. The objection that the 'logic' of the observation-terms is completely determined by the ostensive rules cannot be raised by Nagel who rejects sense data (88, 122), who asserts quite explicitly that ostensive rules define a 'partial meaning' only (88) and who points out that already commonsense 'involves the use of abstract conceptions' (11). The further objection that the logical structure of the observation language is hardly ever available in the form of explicit principles overlooks that the same is true of theories also. Here too, the explicitly formulated laws mirror only part of the 'logical relations in which the terms stand to one another'; the remainder must be unearthed by attending to the way in which theoretical terms are being used. It is plausible to assume that many theoretical postulates have originally been obtained in this manner (examples: the Archimedian postulate in geometry; the law of excluded middle). In the case of observation languages this applies to *all* postulates—this is the only difference. Finally, there will be *models* for P_O. The most familiar models of course assume things having certain properties, but we need not restrict ourselves to them. We may use models similar to those introduced by Whitehead in the case of geometry, and even more unfamiliar ones. There emerges therefore a complete parallelism, in all the respects examined by Professor Nagel, between the domain of observation and the domain of theories. Let us examine the consequences of this parallelism.

(4) It is often said that the aim of science is the classification and the prediction of observational results, and that the distinctive characteristic of *modern* science is mathematics and the experimental method. This is correct as a first approximation, but one has to add that experiments are used not only for discovering new facts, but also for revealing the detailed structure of facts already known. This is what really distinguishes modern science from its predecessor. The laws and descriptions of the Aristotelian science were supposed to give an account of observationally identifiable traits *exactly as they appeared to the observer* (cf. the famous example of the drop of wine in 10,000 pails of water, *de gen. et corr.*, 328 ^{a}27). The point of view of Galileo and Newton (which was prepared by the ancient atomists) aims at an *analysis* of such traits that first eliminates the contribution of the observer and then tries to show how the remainder results from an interplay of processes which can be isolated only with difficulty, or perhaps not at all (and which will have to be described by theories in Nagel's account). We see already here that the familiar characterisation of science as an enterprise that 'aims to understand . . . observable things by discovering a systematic order in them' (Nagel, p. 78) is still quite Aristotelian, especially if we take it together with the remark that observational laws and observational notions 'have a life of [their] own' (87). And we anticipate that a different account will have to be given of the role of observation and experiment.

A science that is interested in analysis uses theories for describing the 'hidden' processes which constitute the observable traits. Such theories must be rich enough to express the specific features of the *individual* processes, and they must also be capable of giving an account of the manner in which all these processes *collaborate* and bring about the overt traits and observational facts to be analysed. In other words, they must provide statements and laws suitable for expressing such facts. Of course, the statements need not *repeat* what was said about the facts at an earlier time. Nor is it advisable that they should. The earlier formulations will usually contain the gross reactions of the observer and the crude beliefs emerging from them and it is just such beliefs one wants to exclude. All that is needed is that the order introduced into the observable material by the theory be comparable, both in complexity and in effectiveness with the order established by the concepts and laws already in use. The exact form of the order need not be preserved. Thus the explanation of Kepler's first law by the general theory of relativity (Schwarzschild's solution) neither gives us an exact ellipse, nor are the predicates describing the shape those of ordinary physical geometry. They do not even share a core of meaning with the preceding notions.[1] Yet the descriptive apparatus of the explanation can well compete with that of the older astronomy. It is even preferable as it no longer contains the incorrect assumption of a spatio-temporal structure that is independent of the state of motion of the observer. It follows, then, that a satisfactory theory contains a calculus C^O_T and postulates P^O_T which, though not coinciding with C_O and P_O can yet provide just as effective a logical skeleton for an observation language.

To return to Professor Nagel's account. In this account the ostensive rules are applied to the descriptive terms of C_O; the terms of C_T, C^O_T included, get their interpretation only indirectly, with the help of correspondence rules. What has just been said would seem to suggest that C_T should perhaps be interpreted in a more direct fashion, without the detour through C_O and correspondence rules. There is no reason why the overt procedures which according to Nagel are used to 'fix a definite, if only partial, meaning' for the descriptive terms of C_O should not be applied to the descriptive terms of C^O_T and in this way fix a definite, if only partial, meaning for *them*. The remainder of the theory would receive its content not through additional linkages, added from without, but via the deductive relations between P_T, C_T and P^O_T, C^O_T. C_O, P_O and correspondence rules would be superfluous. For example, one might apply the operations usually employed for giving content to the term 'heavier than' of some commonly understood idiom to the term 'possessing much greater relativistic mass than' of the special, or the general theory of relativity.

This *direct method of interpretation* as we shall call it would seem to reflect much better the features of a science that progresses by analysis than does Nagel's method. It avoids unnecessary duplication of terms and relational systems (C_O with P_O versus C^O_T with P^O_T). It prevents the perennial retention of the errors implicit in the older observational idiom (for example, it prevents the perennial retention of a language that is not Lorentz-invariant; or of a language of mental terms that does not allow for the incorporation of materialistic features). The charge of circularity which Nagel might raise and the related charge that such a decisive change of the observation language may leave 'nothing for the theory to explain' (87) counts little in comparison with such advantages. Besides, it adds another point in favour of the direct method. For the need to change the observation language indicates that there was no explanandum to start with, or at least, that the older experimental laws were not the proper explanandum. What *did* need explanation was how people could believe them—but this is an altogether different matter.

---

[1] Cf. *Journal of Philosophy*, 1965, lxii, 266-274

(5) Nagel's views on the *interpretation* of theories are closely connected with his account of *reduction* which is 'the explanation of a theory, or a set of experimental laws established in one area of inquiry [the so called secondary science] by a theory usually, though not invariably formulated for some other domain [the primary science]' (338). The most striking cases of reduction are those where a 'set of distinctive traits of some subject matter is assimilated with what is patently a set of quite dissimilar traits' (339 f.). In these cases the 'laws of the secondary science' contain terms that do not occur in the theoretical assumptions of the primary discipline' (352). A reduction will now be possible only if 'special hypotheses', connecting the dissimilar terms of both disciplines are used in addition to the principles of the primary science (365; 353 f.). The *existence* of these hypotheses must not be understood as expressing an 'ontological hiatus' (365) or some 'experimental, or even "metaphysical" facts about some allegedly "inherent" . . . *properties* of objects' (369). It rather expresses the '*logical* fact' that the terms of two sets of statements stand in certain relations to each other (369).

The *content* of the connecting hypotheses depends on the context in which a reduction is to be achieved (357). The principles of the primary science may be connected with experimental notions by these hypotheses only: they 'cannot [then] be checked by experiment, but function as coordinating definitions' (356) (case one). On the other hand the principles of the primary science may have been interpreted independently of, and perhaps already before the secondary science. The reduction of thermodynamics to mechanics is an example of this kind. In this case the 'expressions belonging to [both the principles of the primary and the secondary science] possess meanings that are fixed by [their] *own* procedures of explication . . . [and] are intelligible in terms of the rules or habits of usage of the [corresponding] branch of inquiry; and when those expressions are used in that branch of study, they must be understood in the senses associated with them in that branch, whether or not the science has been reduced to some other discipline' (352—a repetition, on a higher level, of the remark about the independent life of experimental laws). Clearly the connecting hypothesis 'will [now] be one that does not hold as a matter of definition, and will not be one for which logical necessity can be claimed' (358) (case two). Most reductions are supposed to be of this kind. Thus the reduction of thermodynamics to mechanics, or of chemistry to contemporary physical theory 'does not wipe out, or transform into something insubstantial, or "merely apparent" the distinctions and types of behaviour which the secondary discipline recognises', nor does the reduction of headache to neurophysiology 'establish a logically necessary connexion between the occurrences of headaches and the occurrence of certain events and processes specified by physics, chemistry, and physiology' (366). The doctrine of emergence, to mention another case to which the above considerations are applied, is on firm ground 'when construed as a thesis concerning the *logical relations* between [statements describing the emerging traits and statements of the quite different theory used to explain their emergence]' (372). And so is organismic biology insofar as it asserts that connecting hypotheses linking biological terms to terms of physics are necessary for a reduction, but not yet available (434).

(6) From these remarks there now emerges the following view of the structure and, possibly, the development of science.

In so far as they can be related to each other the explanatory systems, or the theories of science are arranged in *layers*, or *levels*, starting with an observational level and moving up to levels of increasing abstraction. Each level is self contained. Having methods and procedures of its own (352) it gives rise to an equally self contained domain of knowledge. The layers are 'linked' by 'rules of correspondence' with the observational level (93 ff.) whose laws have 'a life of (their) own, not contingent on the continued life of any particular theory that may explain them' (87). In addition the layers may be connected with each other by 'specific hypotheses' (365; 473). The rules of correspondence give empirical content (90, 166, 300) and they also make possible the explanation of observable phenomena (93, 97) which are the 'ultimate points of departure' for 'scientific thought' (79). Connections *between* layers are necessary if lower levels are to be explained by, or reduced to, the higher levels. As we have seen such explanation neither 'wipe[s] out' nor 'transform[s] into something insubstantial' the level explained (366). It leaves intact the self contained character of each level, and the relative independence of each domain of knowledge.

Now if we take it for granted, as does Nagel, that explanation is essential for science (3); if we assume that it is 'the desire for explanations which generates science' (4); then the above arrangement of layers becomes a model not only of the *end result*, but also of the *development* of our knowledge.

The development of knowledge then consists in the accumulation of facts, and of theoretical layers. Science advances by internal improvement of each level, by addition of new facts on the observational level, by addition of new explanatory systems on top. There may of course be a considerable amount of change preceding the establishment of each single layer. Theories may have been formulated tentatively, and may again have been abandoned. But as soon as a theory is 'established in one area of inquiry' (138) as soon as it starts pervading this area with its methods and its specific terminology, in the very same moment it assumes that independence of procedures and rules of usage which we have reported above. The lower levels then actually become somewhat assimilated to the observational language and their abstract conceptions assume a similar 'life of [their] own'. This, then, are the outlines of what one might call the *layer model of scientific knowledge.*

(7) This ingenious model which tries to combine the empiricist's concern for *facts* with the realisation of the importance of *theories* and which also tries to satisfy the wish for some amount of stability is in agreement neither with actual science, nor with reasonable methodological demands. The reasons for this have already been stated though so far they have been restricted to the relation between theories on the one hand and experimental laws and observation languages on the other. It has been pointed out that Nagel's method involves an unnecessary duplication of relational systems. And it was suggested that the ostensive rules which connect terms with overt operations and traits should be linked to the theories directly, and not via an older observation language and correspondence rules. In this way the older observation language would be abandoned and its abstract concepts replaced by the abstract concepts of the theory. The charge of circularity was answered by showing that the non-circular procedure preserves (at least part of) the structure of the observational laws and thereby removes it from the domain of criticism. This point deserves repetition: If it is the business of science to explain; if explanation 'takes its ultimate point of departure from problems suggested by observing things and events in common experience' (79); if the description of these things is allowed to have 'a life of its own' (87) that can be expressed 'independently of [any] theory' (86); if a change of theory changes only the 'theoretical interpretation' of the experimental laws and observational terms, but leaves unchanged another part of their meaning (87)—then the relations between terms which constitute this 'observational core' will forever be exempted from change. Their stability will be guaranteed not because a detailed examination has found them to be adequate (Nagel admits that empirical laws may be as 'speculative' as theories (80)), but because they have been arranged in such a manner that an examination becomes impossible. They have been turned into veritable *idola fori.* And such *idola* can be criticised only by a procedure that permits us to compare them with other relational systems, for example

with the observation languages which universal theories give on direct interpretation. Such a comparison, though being perhaps 'circular' is thereby shown to have a definite advantage over and above the method described by Nagel. Now in the case of reduction the situation is exactly the same. But it is somewhat easier to present it in an acceptable manner.

Assume for that purpose that we are dealing with the explanation of classical (Newtonian) momentum conservation by the special theory of relativity. Nagel himself regards the relativistic notion of mass as a 'new notion' and he points out that the conservation principles 'must be reformulated in terms of relativistic mass' (111). He gives a brief, though very enlightening account of the manner in which the relativistic concept differs from the classical concept (170). Newtonian mass is an 'intrinsic property' and it is 'additive'. Relativistic mass is a function of the relative velocity, (and it is non-additive). True, he seems to be inclined to restrict the modification to particles moving with great speed (' the mass of a particle *moving with great speed* varies with the velocity, so that the [classical] principle of conservation of momentum does not appear to hold for *such* particles' (111—my italics)—but I do not think that this passage should be interpreted as saying that the dependence on velocity *disappears* at low speeds (though it is of course to be admitted that it can no longer be *observed* at low speeds). Now the first thing we note is that at low speeds, and after omission of effects unlikely to be spotted experimentally, the theory of relativity gives *exactly the same predictions* as does Newton's theory, and we also obtain a 'sentence *similar in syntactical structure* to the standard formulation [of Newtonian momentum conservation]' (358—my italics; the reference is to the reduction of thermodynamics). In the domain in question ($v/c \ll 1$) the theory of relativity is therefore as effective a predictive device as is the older point of view. But of course, the 'similar syntactical structure' has a meaning that is 'unmistakably different from' what Newtonian momentum conservation asserts (358—the reference is again to thermodynamics). This follows from what was said above: the theory of relativity uses the relativistic notion of mass (dependence on v/c; non-additivity) under *all* circumstances, and not only for velocities comparable with the velocity of light although for $v/c \ll 1$ the predictions (i.e.—the *numbers*) resulting from such use will be indistinguishable from the numbers produced with the help of the classical terms. We see, then that the theory of relativity can do everything the classical theory can do; that it can do a few more things; and that it is successful where classical physics fails. And it would seem that we can safely *abandon* the classical point of view and live with relativity instead. Of course this does not mean, as has sometimes been suggested, that we shall abandon the splendid mathematical machinery of classical celestial mechanics and that we shall regard the achievements of this discipline as being of historical interest only. Abandoning classical physics means abandoning the classical *concepts*; the classical *formulae* may be retained.[1]

However this is just the procedure which encounters Nagel's objection. It may be granted by the objection that we have made some progress. But an *explanation* of classical mechanics has not been achieved. Such an explanation demands the continued use of the classical concepts (cf. the argument at bottom of page 357 as well as the presentation in section 5 above). Let us now examine this objection.

The objection is based on the following two assumptions: (i) science proceeds by explanation; (ii) explanation of well established laws is by deduction. From these assumptions it is inferred that there must be ' special hypotheses ' (cf. section 5) connecting the relativistic notions with the classical notions. Using these hypotheses we obtain pairs of theories whose elements are capable of making predictions in a certain domain, one of them being capable of making predictions in other domains also. The first question that arises is why science should be cluttered up with double systems of this kind when one half can do the job of both. The second question concerns the connecting hypotheses themselves. Here we must distinguish what we above called case one and case two (section 5). Case one means now explaining relativistic mass in terms of classical mass—which certainly is not in accordance with Einstein's intentions (though it may be in accordance with the original intentions of Lorentz). Another example will make the situation even clearer: the phlogiston theory, especially in the form in which it was developed by Stahl and the new ideas of calcination and oxidation form a perfect pair for reduction. It does not matter that the phlogiston theory has now been abandoned—quite the contrary, this is an argument in favour of our procedure. What does matter is that at the time in question it was well established and, because of Stahl's particular prejudices, thoroughly observational. The method discussed at the present moment would amount to *retaining* the notion of phlogiston and the theoretical context surrounding it (' the expressions belonging to a science possess meanings that are fixed by its own procedures . . . whether or not the science has been reduced to some other discipline ' (352)), interpreting all the new notions by reference to it. It would amount to saying that new conditions for the production and transfer *of phlogiston* had been discovered (cf. the similar locution in the case of headache, p. 366). However, we prefer to say that the notion of phlogiston had turned out to be inadequate and that we had found reasons to abandon it.

Case two assumes that the special hypotheses connecting classical terms and relativistic terms are empirical statements asserting ' that the occurrence of the state of affairs signified by a certain theoretical expression . . . in the primary science is a sufficient (or a necessary and sufficient) condition for the state of affairs designated' by an expression in the secondary science (354). In the present case they would assert that whenever a certain relativistic state of affairs is realised, intrinsic mass that is additive (cf. p. 170) is present also. But the theory of relativity *roundly denies* the existence of intrinsic masses and therefore contradicts the hypothesis. Again, the method recommended by Nagel leads to undesirable consequences.

---

[1] Though it seems to me that too much emphasis even on the formulae may prevent us from predicting relativistic effects where we do not expect them.

Now this method follows from assumptions (i) and (ii) concerning explanation. The fact that it leads to undesirable consequences shows that at least one of them is in need of change: we shall either have to admit that science does not proceed by explanation, or that explanation (or reduction) is not by deduction. As far as I am concerned the problem as to which way to choose has no interest whatever. If we know how science proceeds, if we know what moves can be recommended as reasonable, then we know all we need to know. The question as to how these reasonable moves should be *described* is a problem for the publicity experts and should perhaps best be settled by introducing a new terminology that is not as loaded as is the term ' to explain '.

(8) Our argument is still incomplete. Nagel assumes that the laws of the secondary science have been ' established in some area of inquiry ' (338) so that they agree with the experimental results in this area. Now while classical mechanics satisfies this condition (the area of inquiry being defined, relative to special relativity, by ($v/c \ll 1$) it is refuted by experiments outside (111). The same is true of the example which Nagel himself uses, viz. the reduction of thermodynamics to mechanics. Here, too, there exist experimental results (Brownian motion) which contradict the secondary science. Moreover the existence of such common phenomena as noise would seem to show that the domain of success of the secondary science is really quite small.[1] Now in order to dispel the impression that my argument against Nagel succeeds only because he has chosen his examples badly and that it depends on whether or not the secondary science is adequate I shall assume in the present section that it is and that no exception has been found. Can Nagel's theory of reduction be retained under these quite unrealistically ideal circumstances? I think it cannot. The reasons are as follows.

One of the most important tasks of science is to *test* the laws and the theories it introduces for the purpose of explanation and analysis. This cannot always be achieved by a direct comparison with facts. Some of the most decisive refuting instances for a theory T are produced with the help of *alternative theories* T′ which contradict T, repeat the success of T, make predictions in new domains at least some of which are then confirmed by phenomena P. We may say in such cases that T has been indirectly refuted by P. The refutation of thermodynamics by Brownian motion is of exactly this kind. Moreover, it can be shown in this case as well as in others that the alternative T′ was *essential,* as a direct refutation of T by P (of thermodynamics by Brownian motion) is excluded by physical laws.[1] Now pairs of theories such as T and T′ are ideal candidates for reduction, and this according to Professor Nagel's own criteria. In reducing T to T′ Nagel would interpret them in a manner that removes the inconsistency (this follows from his general views on reduction as well as from his own example, the reduction of thermodynamics to mechanics). We have shown above that such interpretations frequently run counter to the intentions of the proponents of T and T′. We have also mentioned other disadvantages. We now see that the procedure lowers the empirical content of our knowledge by eliminating valuable tests. It is therefore consistent neither with actual science, nor with reasonable methodology.

---

[1] One should also consider that the existence of a finite specific heat for all objects leads directly to the kinetic theory.

In the meantime I have been informed by Professor Smart that Professor Nagel is prepared to admit the inconsistency of many theories he regards as candidates for reduction and that he suggests construing the connecting hypotheses as probability hypotheses. This does not at all improve things. First, because the remark concerning the unnecessary multiplication of explanatory systems still remains in force. Secondly, the theory of relativity (to mention only one example) excludes intrinsic mass not only with high probability, *but always* (see the above argument in section 7, case 2). Thirdly a weakening of the inconsistency between competing theories would reduce the critical ability of both and must therefore be avoided. No amount of small scale adaptation will do. Nagel's theory of explanation must be altogether rejected.

(9) We now come to the last item on our list, the question of the cognitive status of theories. It is impossible to treat here in detail all the arguments and interesting observations which Nagel presents on this problem. I only want to make two points. First, Nagel points out that 'a defender of either view [realism or instrumentalism] can not only cite eminent authority to support his position; with a little dialectical ingenuity he can usually remove the sting from apparently grave objections'. And he concludes that 'once both positions are so stated that each can meet the *prima facie* difficulties it faces, the question as to which of them is the "correct position" has only terminological interest' (141). 'in brief, the opposition between these views is a conflict over preferred modes of speech' (152).

Now it is certainly correct that much of the contemporary discussion of the issue has precisely the character described by Nagel. It is also obvious that the addition of further *ad hoc* hypotheses may remove the still remaining differences. But when comparing philosophical theories, or physical theories, we should not use their degenerate forms but those formulations of them which contain only a minimum of adaptive measures. And there are versions of realism and instrumentalism which satisfy this demand. The instrumentalist version of the idea of the motion of the earth, to take one of the most important examples, was supported by the argument that a real motion of the earth would contradict the highly confirmed laws of the Aristotelian science of dynamics. The instrumentalistic version of Newton's theory of gravitation was already defended by Newton himself on the grounds that experience had shown matter to be inert so that it could not be regarded as a seat of forces. A realistic interpretation of the quantum theory is rejected on the grounds that it is inconsistent with the conversation laws and the laws of interference. These are strong reasons and they can be removed only by arguments showing that it is desirable to introduce theories which contradict already existing laws. In the last section we have indicated that such arguments can be provided. This removes one of the most serious historical objections to realism.

---

[1] For details of my essays in vol. ii of the *Delaware Studies for the Philosophy of Science* and in vol. ii of the *Boston Studies for the Philosophy of Science* (the latter being a volume in honour of Professor Philipp Frank).

The second point I want to make is connected with the usual debates concerning the distinction between observation terms and theoretical terms. As has been reported in section 2 Nagel objects to regarding the distinction as spurious because it is vague (83). This is certainly a valid point unless the *raison d'être* of the distinction consists in the role it plays in arguments about matters which do *not* admit of gradual transitions. In this case proof of the vagueness of the distinction is proof of its irrelevance. Now one of the most important reasons for making the distinction (and *the* reason for a thinker like Berkeley) is that one wants to know what is *established*, and what therefore can be *known* to exist, and what is hypothetical only. We take it for granted that observable things can exist whereas some would claim that ascribing existence to theoretical entities is not only false, but a logical blunder. Now the distinction between existence and non-existence certainly is not a gradual one. Admitting that observability is a vague concept is therefore tantamount to admitting that the problem as to what exists and what does not is independent of it, that it is a metaphysical problem, and that the question as to what can be observed and what cannot be observed is comparatively unimportant. Exactly the same result is obtained by attending to what has been said in section 3 of the present review.

In this section we pointed out that observation languages which are not sense data languages, which contain abstract notions, and whose use amounts to asserting certain relations can be dealt with along exactly the same lines as theories. There will be a calculus, C_O with postulates P_O and there will also be models. It follows, then, that even the greatest success of the observation language 'may not be assessed as sufficient for asserting the physical existence of the various elements in the substantive model in terms of which the [observation language] is formulated' (117, applied to theories). Not even the existence of observable events is therefore assured and the 'problem of the cognitive status of theories' expands into the problem of the cognitive status of *all* notions of our language. Nagel's book contains hardly a hint concerning this more general problem and this despite the fact that his own analysis of theories and his rejection of sense data leads straight to it. This is true of other accounts also: the ontological status of observation terms is a matter that is hardly examined today.[1] This certainly is a step back from the sophistication that had already been reached by Kant and an indication that the belief in sense data, though denied in *words* still seems to have a strong influence upon the *deeds* of almost all contemporary philosophers of science.

(10) The topics dealt with so far fill hardly one-third of Nagel's book. There are chapters on the quantum theory, on the science of mechanics, and there is a large amount of argument dealing with the social sciences. These arguments are very instructive and I am sure that everyone reading the book will greatly profit from examining them. The reason why I have concentrated on the chapters of the book dealing with more general matters is that the opinions expressed there are widespread, that they are presented in great detail, and supported with better arguments than I have found elsewhere.

PAUL K. FEYERABEND
University of California, Berkeley

The Basic Laws of Arithmetic. By Gottleb Frege. Translated and edited, with an introduction, by Montgomery Furth.
University of California Press, 1964. Pp. lxiii+119. $5.00.
A Study of Frege. By Jeremy D. B. Walker.
Basil Blackwell Oxford, 1965. Pp. xiv+201. 30*s*.

THE profound influence of Frege's work on the development of formal logic and philosophy of mathematics is now well known. However, his philosophical writings, which, unlike his technical achievements, are still of more than historical interest, appear to be comparatively little known or ill-understood, at least by philosophers who are not primarily interested in the foundations of mathematics. This is a great pity, for Frege's thought is precise, penetrating, subtle and often relevant to philosophical problems which are still under discussion. Perhaps a wider recognition of his importance will be hastened by the translation of more of his works into English, and, it is to be hoped, by the publication of papers reported to be still in the possession of his literary executors.

---

[1] An exception is Carnap's article 'Empiricism, Semantics, and Ontology'.

Mr Furth's translation of parts of Vol. I. of *Grundgesetze der Arithmetic* is a valuable supplement to what is already available in English, though some parts overlap with papers in the collection edited by Peter Geach and Max Black. Mr Furth offers us the Preface (for some reason he translates ' Vorwort ' as ' Introduction '), the Introduction (' Einleitung '), and the first 52 sections of vol. i. Two appendices are added, the first including parts of §§ 54, 55 and 91 of vol. i. These are required for a full understanding of the second appendix which is a complete translation of Frege's Appendix to vol. ii, in which he discussed Russell's paradox. In all this Frege's own symbols are reproduced, except that script letters have been used instead of Gothic. A large portion of the Preface, the first seven sections of vol. i, and a slightly ' doctored ' version of the appendix on Russell's paradox (without Frege's symbols) are already available in the Geach and Black collection. The newly translated passages include a criticism of attempts to bring psychological concepts to bear on logical problems, and an attack on Idealism (Preface, pp. xv-xxiv). In addition there are Frege's detailed explanations of several of his symbols (e.g. the *wertverlauf* and definite description operators), his typographical distinctions between different sorts of variables, general remarks on definitions and *bedeutung* (translated as ' denotation '), informal discussion of his axioms and principles of inference and finally some examples of derivations within the system.

The book will not be easy for non-mathematical philosophers to read, but there are passages translated here for the first time which should help to clarify some of Frege's philosophical ideas, such as the notion of ' bedeutung ' especially as applied to function-signs and concept-signs. Thus, a concept-sign (i.e. a sentence from which a name has been removed) has a *bedeutung*, which Frege calls a ' concept ', if and only if every insertion of a successfully referring name into the gap yields a sentence corresponding to a definite truth-value (§ 29). This suggests, as pointed out in the translator's introduction, that if we treat ' has a bedeutung ' as an indissoluble metalinguistic unit, and then interpret all Frege's talk about concepts as being a shorthand for talk about concept-expressions which ' have a bedeutung ', then there is no need to get bogged down in obscure discussions of what kind of ' entity ' a concept is supposed to be. Perhaps Frege's frequent reminders that he is speaking metaphorically in much of what he says about functions and concepts support this suggestion. Thus, without resorting to any obscure metaphysics we can preserve the elegance and comprehensiveness of Frege's theory of sense and bedeutung, according to which the logical features of such varied signs as names, definite descriptions, concept-words, mathematical and non-mathematical function-signs (e.g. ' the father of . . .'), sentences, sentential connectives, and second-level function-signs such as quantifiers, can all be summed up under the principle that if two or more of these signs are combined correctly then the resulting sign (i) will have a sense if and only if the components do, (ii) will have a bedeutung if and only if the components do, (iii) the sense of the whole being fully determined by the senses of the components, and (iv) the bedeutung of the whole being fully determined by those of the components. (Perhaps it would be well to remind reader that this applies to sentences only in so far as they are stripped of their *asserting* role, which Frege assigns to a separate symbol.) The beauty and power of this theory are clearly brought out in Mr Furth's introduction and appear not yet to have been fully appreciated. They justify Frege's decision to construe sentences as naming, or denoting, truth-values. As it stands, the theory will not work in connection with a language which, unlike Frege's symbolic script, includes some concepts which are only partly determinate. This sort of limitation, arising from his pre-occupation with the foundations of mathematics, may partly explain why the theory is not better known and understood. I think such limitations can be removed, but this is not the place to show how.

Mr Furth's fifty-two page introduction is sure to be useful to those who have difficulty in understanding some of Frege's more involved and obscure discussions, both within *Grundgesetze* and elsewhere. There are eight sections, of which the following are most important: § 3, a discussion of the distinction between *sinn* and *bedeutung*, and the application of the distinction to names (complete expressions) and function-signs (incomplete expressions). There is a lucid exposition of Frege's reasons for treating complete sentences as names. Section 4 contains an explanation of Frege's distinction between function and object, including further remarks on the sense in which incomplete signs *denote* functions, and a description of the role of quantification in Frege's system. Section 5 explains the use of Roman letters as free variables. Section 6 gives a very precise account of Frege's 'wertverlauf' operator, stressing the need to distinguish his notion of the course-of-values of a function from the modern notion of a function-in-extension, or class of ordered pairs. The difference between Frege's notion of the extension of a concept and the ordinary notion of an extension is rightly stressed. A detailed analysis of some of Frege's remarks then explodes the myth that he thought of courses-of-values as *denotations* and functions as *senses* of function-signs. This section also comments on Frege's attempt to get round Russell's paradox. Section 7 explains Frege's assertion sign and raises some objections to his view of sentences as denoting truth-values. The rest of the Introduction consists of general and historical remarks, together with the usual translator's notes. All of it is clear, concise, and, in my view, mostly correct about Frege's intentions.

Mr Walker's book on Frege covers a wider range of topics and goes into them in more detail. Further, it has the advantage of an index. However, on all other points it is much inferior to Mr Furth's introduction, for it is often too vague and does not bring out the central aspects of Frege's theory clearly enough. One feels the author struggling to reach a clear understanding of Frege's work, but not quite getting there. This might have made it useful for students to read and criticise, if there had been more references to particular pages of Frege's writings, and if there had not been too many bad examples set in it. Here are some specimens: p. 1: 'The best way to elucidate the notion [of a function] is to make clear exactly what state of affairs is described by the equation '$y = x^2$'.' But such an equation describes *no* state of affairs, unless the variables are interpreted as bound, in which case it is unsuitable for elucidating the notion of a function. Page 17 (cf. p. 51): 'Frege insists as strongly as Russell that we are only justified in using expressions which have the form of singular definite descriptions, e.g. "The so and so", if we know that at least one thing is so and so, namely if we know the truth of some existentially quantified statement....' This is a sloppy description of Russell's views, and is also misleading concerning Frege, since he said this only about 'scientific' contexts, in which speakers are concerned about the truth or falsity of what they say, and he mentions reported speech and statements about beliefs, thoughts, etc., as counter-instances. Page 30: 'For truth and falisty are notions which operate in definite ways; the ordinary truth-tables of propositional logic contain the ways in which they operate... when we are concerned with validity of proof and correctness of inference, we must see that both premises and conclusions conform to the laws of truth, namely the conditions implicit in the truth-tables.' No comment! Page 81: 'Replacement of a component [of an expression] by another with the same sense but a different reference would not alter the sense of the whole, but would alter its reference ("On Sense and Reference", p. 62).' (The page reference is to the Geach and Black collection.) For Frege, two expressions with the same *sinn* must have the same *bedeutung*, for the *sinn* is what ensures that the expression has this bedeutung rather than that one. It is clear from the context that I have not merely quoted a misprint, with 'sense' and 'reference' transposed. Similar remarks occur elsewhere, e.g. on p. 126. The next two quotations contradict the previous one, but are still inaccurate p. 78: 'For an object to become a sign is to acquire a sense, and to be given a reference is the same as both.' And p. 82: 'It seems then that knowing what the sense of the name 'Aristotle' means is more or less the same as knowing what its reference is.' But Frege makes it clear in 'On Sense and Reference' that one can know the sense of a name without even knowing whether there *is* a reference (bedeutung). (Further, do senses *mean* anything?) There are other topics on which Mr Walker seems to be confused. Page 107: 'Finally complete sentences are taken as functions, since like arithmetical equations they state facts.' Here we find (1) confusion between function and function-sign (2) apparent ignorance of Frege's insistence that only *incomplete* signs can be function-signs, and (3) a repetition of the confusion shown in the quotation from p. 1. On p. 118 Mr Walker claims that only in terms of a distinction between positive and negative thoughts can we understand the role of the negation sign as applied to sentences with unknown truth-value. Compare this with what Frege says in 'Negation' (especially pp. 125 and 130 in the Geach and Black edition). Finally, on p. 180 we find Mr Walker asserting that in talking about a function's value-range (*wertverlauf*) Frege was not talking about the arguments of the function but about its values.

These specimens (and more could have been given) should make it clear to anyone familiar with Frege's work how unreliable Mr Walker is as an expositor. Perhaps he has the excuse of youth and inexperience, but some of the mistakes should never have got past a publisher's reader. There are signs, however, that given more time and some helpful criticism Mr Walker could have written a much better book, for many of his mistakes are implicitly corrected in other parts of the book, and he has usefully collected together many of Frege's scattered remarks on topics of interest (though without giving enough detailed references). The book ranges over so much of Frege's thought, and raises so many interesting questions (e.g. did his defined terms always have the same sense as well as the same reference as the definiens?) that readers already familiar with Frege's work may find reading it a useful and stimulating exercise once they have got over being infuriated.

Aaron Sloman
University of Sussex

Philosophical Problems of Space and Time. By Adolf Grünbaum.
Routledge and Kegan Paul, London, 1964. Pp. 448.

Beginning with the consideration of the problems that led to the theory of relativity, there has been a steadily growing realisation that our basic concepts of space and time involve a very pervasive background of tacit assumptions concerning the structure of the world. Such assumptions originate largely in that domain of thought which is generally called philosophical, rather than in the pragmatic efforts to deal with the observational facts of physics, more or less close to the level at which they are directly presented in experiment. Nevertheless, these philosophical assumptions have an enormous though almost unknown range of influence on how our more pragmatic observations are interpreted, and how the corresponding research activities are organised, directed, and shaped, in such a way that we inquire mainly into questions that are deemed to be significant because they play a part in our tacit philosophical point of view. This book, which deals with a very broad range of basic philosophical problems of space and time, is therefore very pertinent especially to the present phase of development of physics, in which our philosophical notions about space and time are being called into question in the fields of relativity, cosmology, quantum theory and theory of elementary particles.

CALCUTTA CENTRAL LIBRARY

The central question around which the book is organised is that of whether the metrical properties of space and time are objectively existent 'ontological' realities, or whether they are conventions, established because they are useful in the expression and organisation of a broad range of facts of experience. The author takes the view that these properties are conventional However, he does not thereby adopt a pragmatist, positivist, idealist, or solipsist view that the conventional character of the metrical properties of space and time reflects *only* the activity of the mind which is thinking about these properties, or that it is *only* a useful instrument for dealing with practical problems. Rather, his view is that it is an objective fact that these properties are conventional, and can be chosen for convenience in a wide variety of ways, just as, for example, it is an objective fact that on a grid of co-ordinates, the direction that is *called* X is also a matter of conventional choice, and is not a basic ontological feature of the systems whose movements are being described in terms of these co-ordinates. By thus showing the falsity of the assumption that certain actually conventional properties are ontological realities, the author argues, at least by implication, that our minds will be freed to see more clearly what are really the basic ontological elements in space and time. So he is not a *conventionalist* who wants to try to reduce all knowledge to conventions, but rather, his aim is to remove what he regards as a deep kind of confusion between conventional properties and properties which represent objective realities that transcend mere conventions.

Two principle questions arise on this score. Firstly, is our measure of length of space and time a convention? The author argues, fairly convincingly, that as long as we accept continuous space and time as the basic ontological reality, then these must be taken as conventional. Metric would cease to be conventional only if space and time were discovered to have a discrete character, which would imply an inherent 'natural' measure. But in terms of the notion of continuity, there is a wide variety of possible conventions for metric, leading to equivalent observational results, which can be distinguished only by the fact that some conventions lead to what seem to be 'simpler', more 'beautiful', more 'elegant', or more 'natural' theories than others. Indeed, as has in fact been noted by Bondi, the choice between Einstein's ideas and Hoyle's ideas in this general subject must at least for the foreseeable feature, depend largely on such essentially aesthetic criteria rather than on factual tests. And even if there were factual tests favouring a particular theory, these could at most show that theories based on certain conventions had a greater explanatory and predictive power, in the sense that on the basis of 'simpler' hypotheses they related a wide range of data that other theories might relate with more complicated and less natural (i.e. more arbitrary) kinds of assumptions.

The second important place where conventions come in is on the definition of simultaneity. Here, the author makes it very clear that the question of whether simultaneity is conventional or not is itself a *factual* one, related to the ontological character of space and time. Thus, if Newtonian assumptions were true, then it would in principle be possible to send a signal with unlimited velocity from one point to another and back, so that each point would be in immediate two-way contact with all others. If such contact were actually possible, then all points that are so related would be simultaneous in a simple 'ontological' sense, corresponding in essence to our intuitive notion that all that is perceived at a given moment called 'now' is coexistent and capable of immediate two-way contact. But such a notion of simultaneity leads to the Galilean transformation between frames corresponding to observers moving at different speeds, and therefore to the conclusion that such observers must obtain different measured speeds of light. As shown by Einstein, once the fact is established that all observers obtain the same *measured* speed of light, then it follows that each observer is defining the time of occurrence of a distant event by the *convention* that the time taken for light to reach him from a distance, r, is $t = (r/c)$. Different observers, following the same convention, each in relationship to his own observations, will then be seen to ascribe the property of simultaneity to different sets of events. So it is a fact that real space time is such that simultaneity is a convention. What is *not* a convention, but a basic 'ontological' reality is the *relationship* that two events A and B are (or can be) connected by a light signal, which is ascribed invariantly by all observers, independently of their state of motion. In this way, we see that the discovery that the property of simultaneity is actually conventional reveals an important ontological feature of space and time, which is very different from what is suggested in common experience.

The author analyses these questions very extensively, relating them in great detail to the works of other philosophers. In addition, he gives an extensive discussion of various philosophical ideas on continuity, causality, isotropy and anisotropy of time, the asymmetry of retrodictability and predictability, the flow of time, as well as criticising various critiques of Einstein, including the ideas of Whittaker, E. A. Milne, Whitehead, and others.

On the whole, the book seems to make a useful and timely contribution to the subject. However, certain features of its presentation interfere seriously with its effectiveness and clarity. Firstly, the book is unnecessarily long by a factor of at least two, if not more. The author treats too broad a range of subjects, and in addition, each subject is treated with a confusingly extensive development of details, in which the reader tends to lose himself as he forgets what the main point is. Part of this detail is the result of an extremely polemical attitude, in which the author's interest seems as much to be to demolish his opponents as it is to present what he has to say simply, clearly and factually. If the book had left out such irrelevant details, the remaining core of cogent material, presented from what seems to be a basically very useful and illuminating point of view, would very probably have come across to the reader much more forcefully and effectively than it does. Nevertheless, in spite of these defects of presentation, the book is well worth reading for anyone who wants to acquaint himself with basic philosophical problems in our space and time concepts, especially those having to do with the question of defining the extent to which these concepts have a conventional character.

DAVID BOHM
Birkbeck College, London

'Towards an Historiography of Science', *History and Theory, Studies in the Philosophy of History*, Beiheft 2. By Joseph Agassi.
The Hague, 1963. Pp. viii+117. 14s. 6d.

To readers acquainted with the current literature of the history of science, Joseph Agassi's 'Towards an Historiography of Science' exhibits one of the most wide ranging and perceptive minds now active in the field. No lesser evaluation sufficiently accentuates the frequent brilliance of the historical formulations and historiographic *apercus* from which much of his monograph is compounded. Yet what results is principally an exhibition of virtuosity. To perceive great talents in Agassi's monograph is simultaneously to be frustrated and depressed by the relative triviality of their application. The historical profession, and perhaps also the philosophical, has far more to expect from this source.

Most writing on the history of science is, Agassi argues, vitiated by at least implicit adherence to either of two inadequate philosophies of science. The first of these, inductivism, has until recently been dominant, and Agassi is at his best in describing, illustrating, and prescribing for its shortcomings. The inductivist, he suggests, equates scientific knowledge with the laws and theories gathered in up-to-date science textbooks. These nuggets of knowledge, according to inductivism, have been discovered by the contemplative manipulation of raw data previously assembled by unbiased observers and experimenters. Perhaps no philosopher ever held a position so simplistic, but there have surely been historians who wrote from one sufficiently like it to justify Agassi's stance. And their results, as Agassi amply demonstrates in his text and discursive footnotes, have been a shambles. Believing that science is created in the analytic units required by textbook pedagogy, Agassi's inductivist is unable to isolate the 'real' contribution for which his heroes should be admired. Convinced that error is ultimately attributable to bad method (else the facts would speak for themselves), he must either suppress the obvious or transform to a schizophrenic the scientist who at once defended false theories and contributed to the growth of truth.

On these and a series of related points Agassi has perceptive and persuasive things to say. (He suggests, for example, that an excessive concern with method causes the inductivist-historian to overestimate the role of technological, and thus of economic, factors in the overall development of science.) Furthermore, Agassi has remedies for the evils he describes (some of them already put to practice by the conventionalist followers of Duhem, who constitute the second of the groups he is concerned to criticise). The historian must, Agassi urges, examine discoveries within the context of ideas from which they arose. Those ideas, however out-of-date and wrong-headed they may seem today, were once both viable and exciting. And among them are to be found both the main sources of, and the principal barriers to, the discoveries which, the inductivist supposes, emerged from fact alone. Add that Agassi emphasises the need to examine more closely the roles of controversy, problems, and criticism in the history of science and the main thrust of his argument will be apparent. What must still be underscored, however, is the numerous historical examples, miniature case studies, through which that argument is at once illuminated and documented. Ampère, Black, Dalton, Laplace, Lavoisier, Oersted, and Priestley, all figure prominently in Agassi's monograph; about each of them, and others besides, he has suggestive things to say.

Except to the sometimes outrageously selected targets at which he aims his darts, Agassi's historiography is stimulating reading, often both cogent and perceptive. But one may applaud his conclusions, even enthusiastically, and still ask why a first-rate mind bothered to set them down. To whom is the monograph addressed? What did Agassi hope to achieve through its publication? His tone suggests that a principal aim was reform in the contemporary professional practice of history of science. But for this task his monograph is dreadfully misdesigned. To castigate contemporary practitioners by analysing the historiography of James Jeans ('a brilliant writer and an important figure', p. 4), H. E. Roscoe ('his authoritative *John Dalton and the Rise of Modern Chemistry* (London, 1895)', pp. 79 f.), or Max von Laue (p. 2) is laughable. Equally irrelevant, though not laughable, are the repeated disquisitions on George Sarton. Agassi's own footnotes inadvertently demonstrate that contemporary historians reject Sarton's concept of the subject while retaining an often deep admiration for his dedication and scholarship. And even among past giants like Sarton, Agassi is strangely selective. Where, for example, are Brunschvicg, Dijksterhuis, Lange, Lasswitz, Metzger, Meyerson, Tannery, and Wohlwill? Agassi cites a few of them for historical documentation, but discusses none of their work. Yet they represent at least two major historiographic traditions which his monograph scarcely touches.

Nor is Agassi's discussion of contemporary historians more satisfactory. He repeatedly cites 'Koyré's school' as exemplary. But there is scarcely an explicit discussion of its work even though a majority of professionals now under forty-five are probably members. And on the rare occasions when he does deal with a modern, Agassi often chooses men who would agree almost entirely with his historiographic theses (there are many of them), borrows heavily from their work, and then castigates them for missing points that he has seen there himself. No inductivist has been more guilty of attributing all mistakes of fact or interpretation to errors of method. The historian who fails to find in his material what Agassi discovers there is quickly labelled an inductivist, conventionalist, or some bastard mixture of the two. Contemporary historians of science can profit from criticism, but Agassi's blunderbuss is aimed at the wrong targets and too often misses even them.

His own historical examples are, of course, far more likely to be of use. When he turns his attention to the stuff from which history is made, Agassi is invariably perceptive, often brilliant, and sometimes original. But even his best examples remain, as he insists, speculations that are still to be worked out. And if his own 'first maxim of enlightened or broadminded historiography' is widely adopted, one wonders who will be left to perform that task. 'Any interesting or stimulating story is good, and should count as history if . . . it does not *often* violate factual information *easily* accessible to its author' (p. 47, my italics). Agassi's practice is consistent, but Clio, like Urania and her much younger sisters, sets higher standards of relevance and of discipline.

The tragedy in all this is that the surface on which Agassi skates with such brilliance covers large problems. Inductivism, if largely rejected by mid-century historians of science, has had gigantic appeal and still retains it in many important arenas. Since that viewpoint is also patently wrong, one would like to account for its longevity and to examine the manner in which it and its influence have developed. Agassi, who is here little concerned with either sociological or developmental problems, implies that it was fathered fully grown by Bacon and has retained the authority of his aegis ever since. But his own historiography makes this account implausible, and it is independently clear that inductivism has a longer and more varied history during which it has derived authority from many sources, the ideology of professional science among them. How are we to understand these developments? Or again, Agassi's historiography, which is very close to my own, suggests the need for important modifications in contemporary philosophy of science. But with these, too, he almost entirely declines to deal. Having demolished inductivism, his critique of conventionalism is superficial, and he suggests only that it be emended with elements borrowed from the writings of Sir Karl Popper. Perhaps he is right that the job can be done in that way, but it remains to do. If directed to this problem or to one of the other serious issues his monograph evokes, Agassi's next book could be consequential. And a substantial work of history, guided by the intellect which the monograph displays, might be memorable indeed.

Thomas S. Kuhn
Princeton University

Hobbes's System of Ideas: A Study in the Political Significance of Philosophical Theories. By J. W. N. Watkins.
Hutchinson, London, 1965. Pp. 192. 15s.

Mr Watkins's book can be read either as an argument in favour of the proposition that philosophical ideas can imply a political theory, with Hobbes's thought as the ground on which the argument is conducted, or else as an exposition of Hobbes's philosophical and political thought, only incidentally illustrating the thesis that the politics are implied by the philosophy. Different readers will find their own emphasis in the book, but no reader will fail to be impressed by the sureness of Watkins's grasp, by the penetration and subtlety of his understanding of Hobbes, by his pertinacity and rigour. Especially his rigour, which I venture to imagine would have pleased Hobbes himself, who was given to fits of choler whenever he encountered a mind which he judged to be indolent or a form of reasoning which he took to be slack.

Let us deal first with Watkins's argument about philosophical implication. He begins by saying that, *contra* the logical positivists, philosophical ideas *are* possible, defining these as ideas (such as certain of Hobbes's cosmological opinions) which though not empirically testable, conflict with hypotheses which are. Most of Hobbes's fundamental conceptions, such as his materialism, his nominalism, and his determinism, belong to this class of untestable but nonetheless factual statements, and are thus categorised as philosophical ideas. Although I can agree that Hobbes's assertion 'whatever is, is corporeal' is a non-verifiable statement, I am not clear how it conflicts with a statement which is verifiable, but the point is not really important, because whatever the status of Hobbes's materialism as a philosophical idea, there is no doubt that he affirmed it, and the question is, how is his materialism connected to his politics? Clearly the materialism does not entail the political theory in any simple linear connection, as Watkins points out; but, together with his other fundamental conceptions, Hobbes's materialism implies his civil philosophy in the same way, for example, that the law of evidence has important implications for statements made by witnesses in lawcourts, although it does not entail any of those statements. On this question as on a number of others Watkins is pleased to acknowledge the debt he owes to Karl Popper.

In support of his claim that the civil philosophy is implied by the philosophical ideas, Watkins analyses the ideas themselves. He does so crisply and clearly. Also, he draws upon evidence from the history of science and from intellectual history, as well as (though only rarely) from the biography of Hobbes. Inevitably he is drawn into controversy with other Hobbes commentators and it is no small part of the merit of his book that he treats his adversaries with good humour even while he is repudiating their most cherished opinions. It is easy to see why he keeps his temper—he has the best of the argument. Especially is this true of the two collisions in the book which have the greatest impact—the one with Professor Leo Strauss about Hobbes's science, and the other with Professor Howard Warrender about Hobbes's ethics.

Professor Strauss believes that Hobbes's political opinions were formed in the context of the humanist, moralist, essentially pre-scientific tradition, and that when later in his life Hobbes became enamoured of geometry and of Galilean physics he felt himself obliged to recast his political doctrine into the alien mould of a mechanistic theory of nature. The original moral and political outlook was thus deformed, although its essential character was never altered. Watkins has no difficulty in showing that this view is plainly incorrect. Hobbes's earliest political statement, set down in his introduction to the translation of Thucydides, expressed a bias in favour of the study of history as a guide to rational conduct. Under the shaping influence of the new scientific outlook, Hobbes abandoned this 'inductivist' approach, preferring instead a resolutive-compositive method of speculating about civil philosophy which he derived partly from Galileo, partly from Harvey, but more largely from the philosophers and scientists of the school of Padua. In treating of these latter thinkers, Watkins omits the one person whom Hobbes knew personally and to whom he probably owed the greatest debt, namely Berigardus, author of *Circulus Pisanus*. On the other hand we are offered a splendid insight into the way, not hitherto observed, so far as I am aware, in which Hobbes's doctrine of conatus or endeavour, which he used to distinguish between voluntary and involuntary motion, became the source, by a process of inversion, of Leibniz's concept of the monad.

The issue between Watkins and Warrender turns on the meaning of Hobbes's moral utterances. Did Hobbes believe that the laws of nature are absolute moral laws, that they are morally obligatory even in the state of nature before civil society is formed, and that in a commonwealth they oblige the sovereign as much as they do his subjects? Warrender thinks 'yes', Watkins 'no'. Both agree that Hobbes occasionally wrote as if he were giving comfort to the other's point of view, but Watkins's explanation of these apparent discrepancies is more convincing because he fits Hobbes's statements into the context of intellectual history. Taking Hobbes's remarks about the laws of nature, he says, 'to mean that these laws have a moral as distinct from a prudential character would be like taking Copernicus's use of the term "sunrise" to indicate that he had given up his heliocentric hypothesis'. But Hobbes regarded the laws of nature as prudential rather than 'moral', as precepts of reason found out by reasonable men for their own security against the miseries to which their own human natures would otherwise condemn them. Hobbes's laws of nature are, in Kant's sense, hypothetical rather than categorical imperatives, which is to say that they prescribe actions practically rather than morally necessary. Watkins's treatment of this question is illuminating. Indeed virtually everything he has to say about Hobbes is acute. He has written an excellent book, a lively exposition of Hobbes's thought, which is at the same time a powerful argument in favour of the thesis that Hobbes's political theory is implied by his philosophical ideas.

Samuel I. Mintz

The Foundations of Metaphysics in Science. By Errol E. Harris.
Allen & Unwin, London, 1965. Pp. 512. 63*s.*

It has recently been stated with a good deal of plausibility that of all the scientists who have ever lived, 90 per cent. are alive today. I think it could be said with at least equal certainty that of all the philosophers alive today, less than 10 per cent. have any real concern for relating their work to the world view which is emerging from the studies of scientists. This is particularly true in regard to the biological sciences, especially studies on the neurophysiology of behaviour and evolutionary biology. For example, the statement so often made in certain philosophical quarters—that there can be no thought without language strikes any evolutionary biologist who is concerned with the origin of animal behaviour and of systems of animal communication, as quite ludicrous. Many philosophers have not even begun to think in evolutionary terms, not withstanding the fact that it is over one hundred years since the publication of the *Origin of Species*! Similarly, many philosophers who talk glibly and learnedly about perception show a most impressive ignorance of all those recent developments in sensory- and neuro-physiology which should be most relevant to their thinking. Again one may still hear from reputable philosophers, blandly dogmatic remarks on animal nature based solely on the fact that they have once kept a dog or cat—and ignoring the vast wealth of new knowledge on animal behaviour now available. It is therefore, most welcome to find a philosopher, Professor Errol E. Harris, attempting, with outstanding erudition, the kind of philosophical study which so many scientists feel should be one of the prime concerns of philosophy at the present time.

The main thesis of Professor Harris's book is that natural science is, in fact, the empirical foundation of metaphysics. In this large and most impressive volume, for which a companion volume dealing with case-histories of scientific discovery and reasoning is planned, the author attempts single-handed to survey all the sciences from physics and astronomy through the whole of biology culminating in neurophysiology and psychology as they bear on problems of behaviour, consciousness and mentality. No one scientist could criticise adequately every phase of such a work. Neverthe-less, it appears to your reviewer that Professor Harris proves himself an astonishingly reliable guide through these vast fields of knowledge. This is itself, a tremendous achievement and one for which he should earn enduring gratitude.

Professor Harris thinks it possible to show that the scientific evidence at our disposal, much of which however seems to be unknown to, or disregarded by, philosophers, is directly contrary to empiricist presuppositions. He believes it favours an entirely different conception of the world, supporting a radically different philosophy and demanding a conception of logic wholly other than that which, fostered by empiricism and elaborated by symbolic techniques, has gained such widespread prestige at the present day. He considers that by his method a metaphysical theory may perhaps be outlined and a theory of knowledge be approached, which will have some semblance of scientific stability. With such richness of content it is difficult and perhaps, misleading, to pick out particular features and aspects of the study for comment. Personally, I find the author particularly cogent and impressive in his discussion of the present state of biophysics and his understanding of the nature of evolutionary progress and adaptation. The author also shows great acuity in discussing the nature of concepts such as 'consciousness' and 'free will', and in his criticisms of the idea of emergence as it has been put forward in the past. In considering the origin of order in the universe he sees that the pre-requisite for natural selection is an already existing whole not less than a self-replicating protein molecule. This itself requires for its genesis a complex cycle of chemical reactions and this, in its turn, is found to presuppose an already existing organisation of living matter. He says 'ordered totality is thus logically prior to natural selection.' This is, in fact, just the viewpoint to which so many students of evolution are gradually turning today, namely, that the order which is characteristic of living things itself arises and has arisen in nature, in entities which are far below anything which biologists would recognise as 'living'. In his discussions on mentality he comes to the conclusion that consciousness is not a separate entity, but essentially an activity of articulating and integrating distinguishable elements inherent in the already highly organised and integrated open system of the living being.

In conclusion, he finds it obvious that there is nothing in contemporary science which supports, or gives evidence of the truth of, logical atomism or empiricism. On the contrary, he finds that the developing scientific picture shows the universe as a single and complete system, immeasurably and bewilderingly diversified—a unity in and through prodigious multiplicity. Professor Harris states that the aim of science is to produce a theory of the real and that it is his intention as a philosopher to try to develop *ab initio* the metaphysical theory which the discoveries of science demand and for which they provide the empirical grounds. It is of course, clear that individual scientists will find in this great work statements and points of emphasis about their specialities, with which they disagree. Similarly, philosophers will obviously find much to dispute. Nevertheless, that it is a great work is, I think, beyond doubt; and I feel that there are very few philosophers and philosophically-minded scientists who will not find reading it a great stimulus and even an inspiration. If this work ushers in a period in which philosophers of different complexions make real attempts to grapple afresh with some of the fundamental problems of their subject, but with a full knowledge and understanding of the immense explosion of scientific knowledge which we are witnessing, we shall I believe, see the beginnings of great conceptual advances. If this happens much of the credit will be due to the thinking of a small band of philosophers among which Errol Harris takes a high place.

W. H. Thorpe
Cambridge University

Studies in Plato's Metaphysics. Edited by R. E. Allen.
Routledge and Kegan Paul, London 1965. Pp. xii+452. 70*s.*

Professor Allen's book is a collection of twenty of the most important articles on Plato which have appeared in the English language in the last thirty-five years. Five of them, including for example Ryle's classic paper on the *Parmenides*, date from the 1930s. The rest have appeared since 1953. Most of the pieces are already well known to specialists, who will, however, be grateful for their being made available under a single cover. For the non-specialist this admirably chosen collection provides an excellent introduction to the main questions currently being debated in this field, among them the controversy on the date of the *Timaeus* (where Allen includes the two key papers of Owen and Cherniss) and that on the validity of the arguments in the *Parmenides* (represented here by Vlastos and Geach). On the answers we give to these detailed questions will depend the view we take on what has become the central issue in the interpretation of Plato: did he or did he not radically modify the theory of Forms in later life? The conservative scholar will argue that no such modification was needed since the objections that Plato put into the mouth of Parmenides are wide of the mark: and he will point out that neither the late Platonic dialogues nor Aristotle provides any evidence that Plato attempted a major reconstruction of his ontology to answer those objections. But to this the radical scholar will reply that the *Parmenides* seriously undermines one at least of the two types of metaphor that Plato had used to describe the relationship between Forms and particulars (and that Plato can hardly have been unaware of the fact when he wrote the dialogue); and he will maintain that the conclusion we should draw from the fact that no new or reconstructed metaphysics appears in late Plato is not that he remained satisfied with his earlier ideas, but that he was in serious doubt about how to resolve the difficulties he now realised they involved.

A good deal more work will have to be done before scholars reach a consensus of opinion in this issue. And if this seems surprising or even shocking considering the amount of attention that has always been lavished on Plato, it should be pointed out that this controversy stems in part from a healthy reappraisal of his often underestimated ability as a logician. Meanwhile historians of science may note that the debate on the date of the *Timaeus* affects our interpretation of Plato's role as a natural philosopher, for only if the 'traditional' late date of that dialogue is correct can we be certain that natural philosophy held an important part in Plato's interests to the end of his life. But while we can only be grateful for the care which the epistemology, ontology and theology of that dialogue have been studied in recent years, it is regrettable that the important physical theories it contains continue to be relatively neglected.

G. E. R. Lloyd
Cambridge University

ERRATUM

In the Note and Comment 'A Comment on Miller's New Paradox of Information' by Karl R. Popper in the May 1966 issue of the *Journal*, the penultimate line on p. 68 should read:

(Rcr) If $E^{M}_{a,b}$ is true, then $c(H^{M}, E^{M}_{a,b}) = a/(a+b)$

ANNOUNCEMENT

The Philosophy Department of Wayne State University will start publication in 1967 of *Nous*, a new philosophical quarterly. *Nous* will publish both long essays and short discussions on philosophical problems, regardless of the author's philosophical school or point of view. It will welcome papers that apply formal logical techniques to philosophical questions. *Nous* will publish critical studies of recent books rather than reviews. It will publish symposia. The first issue will contain a symposium on epistemic logic.

Potential contributors are invited to send two copies of their manuscripts to: Professor Hector-Neri Castaneda, Department of Philosophy, Wayne State University, Detroit, Michigan, 48202. Subscriptions may be placed with the publisher: Wayne State University Press, Detroit, Michigan, 48202.

Wittgenstein's Logical Atomism

JAMES GRIFFIN

In this study, now reprinted, of the central philosophical topics of *Tractatus Logico-Philosophicus*, the author's aim is to understand Wittgenstein's programme of analysis. 25*s net*
Oxford Classical and Philosophical Monographs

The Measure of the Universe

A History of Modern Cosmology

J. D. NORTH

'This book comes at a most opportune moment . . . a most illuminating and enjoyable introduction to the subject.' ENDEAVOUR 70*s net*

Infinity

An Essay in Mathematics

JOSÉ A. BENARDETE

An attack on finitism in all its forms, philosophical and mathematical, from Aristotle and Kant to Brouwer and Wittgenstein. 45*s net*

Theophrastus: De Lapidibus

Edited with Introduction, Translation and Commentary by

D. E. EICHHOLZ

This edition, which is largely concerned with textual and philosophical matters, helps to reveal the quality of thought that went into the making of the earliest classical treatise on mineralogy. 4 *text figures* 45*s net*

The Six-Cornered Snowflake

JOHANNES KEPLER

Translated by C. HARDIE

In this lively essay, written in 1611 and never before translated into English, Kepler anticipates the twentieth-century science of forms by asking 'Why Six?' 2 *plates* 6 *text figures* 21*s net*

The Background to Newton's 'Principia'

A Study of Newton's Dynamical Researches in the Years 1664-84

JOHN HERIVEL

'In this long-awaited study Dr Herivel . . . has greatly illuminated the technical and conceptual background essential for a deep historical understanding of the *Principia* . . . the book will almost certainly become a classic in the field.' ENDEAVOUR 5 *plates* 70*s net*

OXFORD UNIVERSITY PRESS

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

SEPTEMBER 1966, VOL. XXVII, No. 1

I. ARTICLES CONTENTS

Happiness and Time, by WLADYSLAW TATARKIEWICZ
Realization and Unconscious Inference, by JOHN WILLIAMSON
Science, Truth, and Ordinary Language, by RICHARD H. SCHLAGEL
Pragmatism and the Outlook of Modern Science, by REUBEN ABEL
An Approach to a Philosophical Anthropology, by RICHARD M. ZANER
Introspection and Incorrigibility, by CHARLES RAFF
Homeostasis, Society, and Values, by RADHAKAMAL MUKERJEE
Indexical Sentences and Cartesian Rationalism, by RICHARD W. PELTZ
About the Problem of Truth, by JAMES E. TOMBERLIN

II. DISCUSSIONS

J. D. Wild on Responsibility, by WILLIAM K. FRANKENA
Reply to Frankena, by JOHN WILD
Reply to Wild, by WILLIAM FRANKENA
The Rhetorical Conception of Truth, by PETER A. CARMICHAEL
On Professor White's Puzzle, by ROBERT W. BEARD

III. REVIEWS, NOTES AND NEWS, AND RECENT PUBLICATIONS

Subscription prices $6.00 per year for individuals; $7.50 per year for institutions. Separate issues, $1 65 and $2 00 respectively. Postage added for back issues only. A circular listing the contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo, Buffalo, New York, 14214

JOURNAL OF THE HISTORY OF IDEAS

A Quarterly devoted to Intellectual History

VOLUME XXVII OCTOBER-DECEMBER 1966 NUMBER 4

TABLE OF CONTENTS

ARTICLES

The History of Probability and the Changing Concept of the Individual . . . *Bayard Rankin*
Perennial Philosophy: From Agostino Steuco to Leibniz . . . *Charles B. Schmitt*
The Estrangement of Celestial Mechanics and Religion . . . *Herbert H. Odom*
Nietzsche as *Politique et Moraliste* . . . *S. J. Colman*

NOTES

The Longevity of the Patriachs: A Topic in the History of Demography . . . *Frank N. Egerton*
Newton's Justification of the Laws of Motion . . . *Margula R. Perl*
Mercurius Politicus under Milton . . . *H. Sylvia Anthony*
"Laissez Faire" in English Classical Economics . . . *Edward R. Kittrell*

REVIEW

Three Faces of Facism by Ernst Nolte . . . *George L. Mosse*

Subscription Rates (outside U.S.A. post paid): $6.50 for one year;
$1.75 for single issue

Payable to

JOURNAL OF THE HISTORY OF IDEAS, INC.
The City College, New York, N.Y. 10031

THE
REVIEW OF METAPHYSICS
A PHILOSOPHICAL QUARTERLY
Edited by RICHARD J. BERNSTEIN

Volume XX, No. 2 *Issue* 78 *December* 1966

ARTICLES

Klaus Hartmann	*On Taking the Transcendental Turn*
Marjorie Grene	*Positionality in the Philosophy of Helmuth Plessner*
Edward H. Madden and Peter H. Hare	*Evil and Unlimited Power*

CRITICAL STUDY

Richard J. Bernstein	*Sellars' Vision of Man-in-the-Universe, II*

DISCUSSIONS: OTHER MINDS

E. M. Zemach	*Sensations, Raw Feels, and Other Minds*
Michael Anthony Slote	*Induction and Other Minds*

BOOKS RECEIVED

Edward A. Reno, Jr. and Staff	*Summaries and Comments*

PHILOSOPHICAL ABSTRACTS

ANNOUNCEMENTS

Annual Subscription $6.00 Student Subscription $4.00

LYMAN-BEECHER HALL, HAVERFORD COLLEGE, HAVERFORD, Pa. 19041 USA

Readers resident in Great Britain may subscribe at a special rate through our British representative, Professor Stephan Körner, Department of Philosophy, The University, Bristol 8, England

VOLUME VIII DECEMBER 1966 No. 2

RATIO

edited in co-operation with
H. B. Acton (London), P. Bernays (Zürich), A. Church (Princeton), H. Cherniss (Princeton), G. Henry-Hermann (Bremen), W. C. Kneale (Oxford), E. Nagel (Columbia), K. R. Popper (London) and J. W. N. Watkins (London)
by
Stephan Körner (Bristol)

ISAAC LEVI	On Potential Surprise
HECTOR-NEIR CASTANEDA	"He": A Study in the Logic of Self-Consciousness
R M. MARTIN	On Theoretical Entities
ROBERT W. BEARD	"The Will to Believe" Revisited

REVIEWS:

RUSH RHEES	*The Philosophy of Wittgenstein* (George Pitcher)
A. C. EWING	*The Case for Rationalism against Modern Philosophical Trends*

Price 15s. net **Annual Subscription: 27s. 6d. net**

BASIL BLACKWELL · OXFORD

MIND

A Quarterly Review of Psychology and Philosophy

edited by Professor **GILBERT RYLE**

CONTENTS FOR ISSUE OF OCTOBER 1966

I.	What has Confirmation to do with Probabilities?	L. JONATHAN COHEN
II.	Explanation and Causality	ARTHUR W. COLLINS
III.	'Real'	JONATHAN BENNETT
IV.	Teleological and Functional Explanations	ROLF GRUNER
V.	The Incommunicability of Content	DANIEL M. TAYLOR
VI.	Is Material Implication Inferentially Harmless?	J. O. NELSON
VII.	Mental Causality	E. M. ADAMS
VIII.	Discussions:	
	Analogy, Verification and other Minds	J. W. MEILAND
	On the Formalisation of Set Theory	J. W. SWANSON
	Criticism and Punishment	JOHN CHARVET
	Iterated Deontic Modalities	RUTH BARCAN MARCUS
	Note on a Doctrine of Frege and Wittgenstein	BEN HAWKINS
	Geach on Supposition Theory	T. K. SCOTT
	Explanation in History	J. F. LING
	The History of the Emotive Theory of Ethics	C. D HARDIE
IX.	Critical Notice:	
	Brand Blanshard, *Reason and Goodness*	R. EDGLEY

NEW BOOKS NOTES

Annual Subscription 25s, post free; single copy 7s 6d, plus postage

Published for MIND ASSOCIATION by

BASIL BLACKWELL 49 BROAD STREET OXFORD

HISTORY OF SCIENCE

An Annual Review of Literature, Research and Teaching

Edited by A. C. CROMBIE and M. A. HOSKIN

CONTENTS OF VOLUME 5, 1966

ARTICLES:

Robert M. Young—Scholarship and the History of the Behavioural Sciences
W. F. Ryan—Science in Medieval Russia: Some reflections on a recent book
Marie Boas Hall—Sources for the History of the Royal Society in the 17th Century
V. A. Eyles—The History of Geology: Suggestions for further research

ESSAY REVIEWS:

R. S. Roberts—The Royal College of Physicians of London in the 16th and 17th Centuries
Walter Pagel—Transition in Chemistry
D. T. Whiteside—Newtonian Dynamics
Laurens Laudan—Method and the Mechanical Philosophy
A. W. Thackray—The Chemistry of History
D. M. Knight—Reduction in Physiology
J. E. McGuire—Intellectual History or Scientific Biography?

BOOK REVIEWS NOTICES OF BOOKS DOCTORAL THESES IN PREPARATION

vi+161 pages, cloth, 42s.

Published by

W. HEFFER & SONS LTD., Cambridge, England

SYNTHESE

An International Journal for Epistemology, Methodology and Philosophy of Science

Editor: JAAKKO HINTIKKA
University of Helsinki and Stanford University

Board of Consulting Editors: David M. Armstrong, Yehoshua Bar-Hillel, Arthur W. Burks, Donald Davidson, Hans Freudenthal, Béla Juhos, Jerrold J. Katz, Imre Lakatos, G. Nuchelmans, Wesley C. Salmon, Erik Stenius, Patrick Suppes, Klemens Szaniawaki, H. Törnebohm.

Contents of Volume 16, No. 2, November 1966

On the Nature and Purpose of Measurement *by Ernest W. Adams.*

Scaling Theory and the Nature of Measurement *by William W. Rozeboom.*

Communications of the International Union of History and Philosophy of Science, Division of Logic, Methodology and Philosophy of Science

Recent Work in Probability and Induction. Review Article *by Isaac Levi.*

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND

Annual Subscription Price Dfl. 48.— post-free

FOUNDATIONS OF LANGUAGE

International Journal of Language and Philosophy

Editorial Board **Morris Halle (MIT), Peter Hartmann (Münster/W), K. Kunjunni Raja (Madras), Benson Mates (Univ. of California), J. F. Staal (Amsterdam), Pieter A. Verburg (Groningen), John W. M. Verhaar (Manila)**

VOLUME II, No. 4 NOVEMBER 1966

Articles Archibald A. Hill: Non-Grammatical Prerequisites.
Anatol Rapoport, Amnon Rapoport, William P. Livant, John Boyd · A Study of Semantic Space
B. K. Matilal: Indian Theorists on the Nature of the Sentence (*vākya*).

Discussion On a Misapprehension of the Status of Theories in Linguistics, by Y. Bar-Hillel.

Reviews By Martha Kneale, Kurt Rudolf Fischer, James D McCawley, Peter Sgall, Jacek Fisiak, Fritz Guttinger and Barbara M. H. Strang.

FOUNDATIONS OF LANGUAGE is published quarterly

Annual subscription price ƒ 42.—($11.75; 84s.)

D. REIDEL PUBLISHING COMPANY / DORDRECHT - HOLLAND

CALCUTTA UNIVERSITY
CENTRAL LIBRARY.

Printed in Great Britain at The University Press, Aberdeen

[illegible]

The British Journal for the Philosophy of Science

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

Cambridge University Press

February 1967

Volume 17 No. 4 15s. net

Brit. J. Phil. Sci., (1966), Volume 17, Number 4, 265–278
Printed in Great Britain

SOME RELATIONS BETWEEN PHILOSOPHICAL AND SCIENTIFIC THEORIES[1]

S. KÖRNER

SINCE the question of the relations between philosophy and science is never very far from the minds of philosophers of science, an occasional attempt at answering it seems permissible and may even be useful. The paper is divided into three parts. First, philosophical and scientific theories are compared by reference to commonsense thought. Secondly, some of the reasonable and unreasonable constraints will be discussed which philosophical theories may impose upon scientific ones, and scientific theories upon philosophical ones. In the third part I try to indicate briefly the relevance of these comparisons to some special problems in the philosophy of science.

I

Commonsense thinking, like philosophical or scientific thinking, proceeds by isolating individuals from their spatio-temporal background, by forming attributes to which it may or may not grant individual status, by applying attributes to individuals or groups of individuals, by discerning relations between attributes and by drawing inferences from statements to statements. The differences between commonsense thinking on the one hand and philosophical and scientific thinking on the other, are differences in degree: the degree to which various purposes of thinking are distinguished from each other and to which thinking is adjusted to these separate purposes; the degree to which standards of correctness and incorrectness are made explicit rather than left to implicitly followed habits; the degree to which precision and exactness is required; and the degree to which the resulting thought is systematised and codified. Scientific and philosophical thinking absorbs much of common sense in more or less modified form and much theoretical thought, after some loosening of its structure, seeps through into common sense.

---

[1] Presidential address delivered to the British Society for the Philosophy of Science on 14th March 1966.

The principal aim of scientific thinking is specialised, systematic, empirically testable knowledge. Scientific specialisation rests on the working assumption that the universe can be divided into separate fields of investigation each of which is free from any but negligible outside influences. This division into separate fields is implemented by restricted ontologies and correspondingly restricted terminologies and is subject to change under pressure of new experience. The systematic character of scientific thinking finds expression in continual attempts at formulation of general statements and their unification in deductive systems or at least systems some of whose statements are connected by more or less extensive deductive chains. Empirical testability in the sciences involves standardised and, where possible, repeatable routines for ascertaining the concordance of theory and experience, ideally the concordance of empirical statements with theories which are in Popper's sense falsifiable.

The original aim of philosophy was comprehensive, systematic and intellectually satisfying knowledge of everything that is capable of being so known. From its beginnings this comprehensiveness of aim did not imply disallowing the claims of common sense and non-philosophical thinking, e.g. in mathematics, to partial knowledge. On the contrary, if its aim was to be brought nearer to fulfilment, the methods and results of non-philosophical thinking had to be clarified and analysed, the relations between different types of non-philosophical thinking exhibited and examined, desiderata and programmes for various types of inquiry set up, and gaps in our knowledge provisionally filled by metaphysical speculation. The ambitious and very likely unattainable character of the original aim of philosophy, has led to a separation of its different functions into two connected branches, on the one hand a philosophy of first order, which—independently of the sciences and sometimes in competition with them—endeavours by methods of its own, e.g. the phenomenological method, to understand the physical, human and social world; on the other hand a philosophy of second order, which inquires into the methods and results, either separately or in their interrelations, of non-philosophical thought about the world. It is particularly in this interest in the interrelations between different fields, that something of the predilection for comprehensive knowledge as the aim of philosophy is still alive. The systematic character of philosophy expresses itself in attempts at explicit formulation of assumptions and presuppositions rather than at deductive unification. As regards the nature of the intellectual satisfaction towards which philosophy is striving, different schools of philosophy offer different accounts of, and criteria for, it if they offer any at all.

The divergence of the aims of scientific and philosophical thinking affects the theories resulting from them. It accounts in particular for some of the limitations of these theories. The deliberate restriction of scientific inquiry to special fields permits the employment of concepts simpler in structure than corresponding concepts not restricted to specific contexts in which features otherwise relevant can be safely neglected. This simplification is based on neglecting what is negligible for some purposes but not for all, and in some contexts but not in all. It has turned out a price worth paying for the sake of the ease with which the simplified concepts can be handled in inference and calculation. This is not the place for a detailed inquiry into how commonsense concepts are replaced by the simplified concepts contained in theories and how concepts which have no commonsense counterparts are formed.[1]

The simplification in question may be either general or specific. The general simplification or idealisation is enforced by the common logical framework of (almost) all scientific theories, i.e. by classical quantification theory and the theory of identity, and the various extensions of elementary logic which are necessary if arithmetic and the theory of real numbers is to be employed. It can be shown both that, and how, (i) quantification-theory enforces the replacement of inexact (extensionally indefinite) empirical concepts by exact ones, (ii) the theory of identity enforces the replacement of the non-transitive relation of empirical indistinguishability by transitive mathematical equality-relations and, (iii) the mathematical theory of measurement enforces the idealisation of physical addition and other physical operations of measurement, etc. It can further be shown how the concept of experienced time is replaced by a mathematical construct of a non-denumerable infinity of pre-existing moments, and how the concepts of effective choice and other (intentional) mental phenomena are 'extensionalised', i.e. excluded from any scientific theory. Specific simplifications or idealisations are not imposed by the logico-mathematical framework of scientific theories, and such idealisations vary consequently from one theory to another. They consist, as it were, in the cutting away of logical connections, as when, e.g. the concept of a physical body which, besides and apart from having a position and momentum, implies also the possession of many other qualities, is replaced by the concept of a configuration of particles, defined as having positions and momenta only.

---

[1] For details see *Experience and Theory* (London, 1966)

Whereas the limitations of scientific theories are rooted in deliberately restricted and simplified ontologies and terminologies and ultimately in specialisation, the limitations of theories which are philosophical, stem from their aim at relative comprehensiveness, an aim which implies their dependence on prior non-philosophical knowledge. Thus if second-order philosophy is to lead to intellectually satisfying theories by the analysis, clarification, qualification and correction of non-philosophical theories, it must be well informed about the facts which these theories seek to systematise, the assumptions made in them, the logical framework in which they are embedded and the range of their applicability. Again, in so far as first-order philosophy investigates phenomena in competition with other inquiries, it must compare their results with its own in order to justify the need for a separate philosophical investigation, where others are, or seem to be available. The phenomenologist philosopher cannot ignore psychology, the linguistic philosopher linguistics, nor, I should think, the moral philosopher anthropology, at least in so far as these inquiries imply or suggest the spuriousness of their philosophical competitors. The external limitation of philosophy by its dependence on prior non-philosophical knowledge has been taken very seriously by most philosophers from Plato onwards—even though non-philosophical competition for the knowledge sought by philosophers is smaller the further one goes back into the past.

One cannot even claim to examine methods or results with which one has no acquaintance, just as acquaintance with them does not commit one to their unqualified acceptance. The ever increasing range of scientific inquiry has perhaps changed the emphasis on the different functions of philosophy, but certainly not the need for it. Whereas before the beginnings of modern science most philosophy was pursued *in lieu* of other inquiries and in order to keep questions so far unanswerable alive (a function which still is and will remain legitimate), an increasing amount of philosophical attention is being given to the examination and evaluation of scientific theories.

2

The above rather rough comparison of philosophical with scientific theories by relating them to commonsense thinking and by pointing to external limitations of philosophical thinking and internal limitations of scientific thinking, throws some light on their interrelations, in particular on the influence which, reasonably or unreasonably, they exert on each other. Among the reasonable influences of philosophical on scientific theories are the reinforcement of the internal limitations of scientific theories by making these explicit, the articulation and modification of principles underlying the construction of scientific theories and the provision of conceptual structures which, though unconnected with sense-experience, are, after more or less radical changes, capable of being so connected.

Scientific self-criticism is naturally directed towards the structure and application of specific scientific theories, rather than towards those features which all such theories or many of them share with each other. It is the task of philosophical criticism to bring these (comparatively) more general features to light and prevent the mistakes which may arise from ignoring them. The simplification of empirical concepts, sketched in the preceding section, i.e. their accommodation to the framework of classical logic and its extensions and the cutting away of some of their logical connections with concepts not needed in specific theories, has the consequence that many of the simplified concepts are no longer strictly, or without qualification, applicable to the substitution-instances of their unsimplified empirical counterparts. This was very clear to Plato who insisted that, e.g. a drawn circle is neither a substitution-instance of 'geometrical circle' nor an element of the class of geometrical circles; and that one must not without qualification identify a drawn circle with a geometrical circle, the concept of a drawn circle with that of a geometrical circle, or the element-class relation with the relation of 'participation' which holds e.g. between drawn circles and the concept of a geometrical circle.

Underlying these qualifications is the principle, as it may be called, of the negligibility of the neglected, i.e. the requirement that if in a certain context an empirical concept and an ideal one are identified, those features which are neglected by the identification should in its context be negligible. The requirement is implemented by the indication of margins of error and by attempts to isolate the contexts of the identification, in particular experimental situations, from all known or suspected external disturbances, such as interference by deliberate human action with the course of a physical process. There are many urgent philosophical occasions for invoking observance of this principle. It must, for example, be invoked to avoid unnecessary paradoxes, such as the paradox of the arrow, which rests on an unqualified identification of perceived continuous motion with a mathematical relation between a linear continuum of spatial points and a linear continuum of temporal moments. It must also be invoked whenever a theory is—in dogmatic, rather than heuristic fashion—interpreted as valid beyond its known range of applicability.

Philosophy does not only exhibit the implicit structure and content of scientific theories, but also the principles implicit in theory construction. Such principles do not necessarily embody detailed methodological advice, which philosophers are in any case not normally competent to give, but are more general in character. In the construction of scientific theories scientists do not only aim at testable systems, but show at different periods different preferences as regards styles of reasoning, paradigms of theories, standards of intelligibility. These preferences depend to some extent on extrascientific convictions, metaphysical, religious, political, etc. Their articulation as regulative principles which formulate standards of intelligibility going beyond deductive systematisation and testability, is one of the tasks of philosophy. Such regulative principles in accordance with their function are best expressed as hypothetical imperatives: If in the construction of a scientific theory you wish to follow the most successful extant theories, the traditional pattern, etc., then . . . If in the construction of a theory you wish to be or remain a good catholic, marxist, humanist, or what have you, then . . . It is by the admixture of extrascientific ideology in regulative principles of theory-construction, that philosophical theories can, beyond reinforcing constraints which are implicit in existing theories, add additional ones. The reasonableness of regulative principles *qua* hypothetical imperatives consists just in their being hypothetical. Their demands are conditional; they do not demand acceptance of the conditions which they formulate.

Applications of the principle of the negligibility of the neglected and regulative principles of the kind described belong to second-order philosophy. The influence of first-order philosophical theories on science consists in the provision of conceptual networks, which, though their applicability is not testable, can yet after more or less radical modifications, be woven into the fabric of scientific theories. There is an analogy between mathematical and speculative philosophical theories which, though not primarily meant for incorporation in scientific theories, have yet been so incorporated. It is often economically reasonable to avail oneself of suitable existing conceptual schemes rather than to construct new ones. The most obvious instances of such transplantation of parts of philosophical theories into scientific ones (for example the case of philosophical and scientific atomic theories) are too well known to need elaboration here.

I now turn to unreasonable constraints imposed by philosophical upon scientific theories. In doing so I shall ignore crude metaphysical dogmatism and attend to some subtler perversions of second-order philosophy. They consist in the unconscious interpretation of features brought to light by the analysis of existing theories or modes of theory-construction as permanent or necessary features of all theories and the consequent rejection as unsound or spurious of any theory which does not possess them. Such restrictions are easily seen to be capricious when highly specific features of highly specialised theories are endowed with alleged permanence or necessity. However, even internal restrictions imposed upon all known scientific theories by the logico-mathematical framework in which they are embedded cannot, therefore, be regarded as unalterable. And the contention that thinking in a different logical frame, imaginable or as yet unimaginable, could no longer be called ' thinking ' is no more than a trivial verbal manoeuvre.

An illustration of the dogmatic misinterpretation of theories is the Kantian thesis that in the commonsense or scientific descriptions of objective experience a certain system of categories is necessarily employed ; as also the thesis which we sometimes find pronounced by recent philosophers of science, that the laws of nature are necessarily statistical. Illustrations of the dogmatic misinterpretation of strategies of theory-construction are the Cartesian thesis that intelligibility must ultimately be reduced to self-evidence, and the Leibnizian principle of sufficient reason. Theories which cannot be adapted to a wider schema conforming to these standards are then regarded as at best convenient stop-gaps without any claim to truth.

Arguments adduced in support of these dogmatic misinterpretations might well after their most famous instance be called ' transcendental deductions '. I know of none which are not either circular arguments or *non sequiturs*. The Kantian transcendental deduction, the invalidity of which is shown by the post-Kantian development of mathematics and physics, is an example of the circular type: First it is shown that ' we ', i.e. the mathematicians, scientists and commonsense thinkers up to Kant's own day, employ certain categories. It is then asserted that we must use them, the assertion being supported by an argument to the effect that we do use them.[1] Descartes' argument in support of his conception of knowledge and understanding is, as has often been observed, a *non-sequitur*, arguing as it does from the alleged self-evidence of an allegedly true synthetic judgment to the assumption of a general concept of self-evidence, the applicability of which to judgments is a sufficient condition of their truth.

I have already pointed out that because of its aim at comprehensiveness philosophy is subject to external constraints imposed upon it by prior non-philosophical knowledge since philosophy must be *au courant* with the state of its subject matter and with those rival enquiries with which it competes for knowledge. Examples come again readily to mind. A philosophy of science which advances the thesis that all laws of physics must be 'causal' in the sense in which the laws of motion are, rests either on ignorance or else on philosophical arguments, which require very close scrutiny especially if put forward as so-called transcendental deductions. Even in so far as philosophy is programmatic (regulative), e.g. by proposing a return to causal explanations, its arguments are likely to become unrealistic if they ignore the state of science the development of which it undertakes to influence.

A clear example of the competition between philosophical and non-philosophical theories is that between philosophical and mathematical theories which have the same mathematical theories for their subject-matter. Ever since the middle of the last century philosophical theses about mathematics had to be corrected in the light of mathematical or metamathematical theses. I have tried to show elsewhere[2] how since the discovery of non-Euclidean geometries, certain doctrines, generally shared, about the nature of mathematics had, as a result of new mathematical knowledge, to be either radically reformulated or corrected. I mean here, in particular, the doctrine of the uniqueness of mathematical reality or intersubjective intuition; of its accessibility in principle, of its unambiguous and exhaustive reflection in conceptual or linguistic formulations. The metamathematical results relevant to these corrections or radical reformulations are, in particular, the proof that the fifth Euclidean postulate is independent of the remaining postulates, the Löwenheim-Skolem theorem, Gödel's incompleteness-theorem, Church's theorem and Cohen's proof of the independence of the continuum-hypothesis.

---

[1] For a detailed criticism which shows not only what I regard as one of Kant's mistakes, but also shows one of the great achievements of the transcendental philosophy see *Zur Kantischen Begründung der Mathematik und der Naturwissenschaften* forthcoming in *Kant Studien* (1966).

[2] 'On the Relevance of Post-Gödelian Mathematics to Philosophy' forthcoming in *Proc. Int. Coll. of Science*, London 1965 (Amsterdam, 1967).

Some of the unreasonable constraints imposed in the name of science upon philosophical theories are, as it were, mirror images of the reasonable constraints imposed by philosophical upon scientific theories. They stem from the internal limitations of scientific theories and from the failure of scientific self-criticism or philosophical criticism to watch over the implementation of the principle of the negligibility of the neglected. Thus the great achievement of a scientific theory in covering a field of enquiry by means of a partial ontology and a corresponding terminology is misused by the claim that the ontology is complete and that any category or concept which is not contained in the partial ontology is empty. Dogmatic behaviourism is a good example of this misuse, in so far as it is based on the assumption that psychology or 'science' has shown that all concepts of mental phenomena are empty, or that they are analysable in terms of concepts of physical phenomena. Psychology or any other science has, of course, shown nothing of the kind, and the assumptions have to be supported by independent philosophical arguments. The alleged emptiness of ' mental concepts ' is usually supported by arguments to the effect that there can be no evidence for the applicability of these concepts, since such evidence would presuppose self-awareness and since the notion of self-awareness leads to an infinite regress. It is, however, easily shown that the regress is harmless. Arguments to the effect that mental concepts are analysable in terms of physical ones, proceed by establishing equivalence between statements involving the former and statements involving the latter, where the equivalence in question is at best only strong enough to establish correlations between mental and physical phenomena—correlations which would be perfectly acceptable to anybody who should insist, like, e.g. Leibniz or Brentano, on the irreducibility of mental phenomena to physical ones.

Dogmatic behaviourism as the doctrine of the spuriousness of mental concepts must be distinguished from methodological behaviourism, i.e. a (regulative) principle which excludes them as not needed from a certain type of inquiry: What is not needed may yet be available and what is available may yet not be needed. The rejection of dogmatic behaviourism is compatible with the acceptance of methodological behaviourism and a conviction of its value as a scientific theory. Dogmatic behaviourism is only one species of scientism which implies that the union of all scientific ontologies is the complete ontology. The rejection of this rather vague and ambiguous doctrine is, of course, quite compatible with the conviction of the value of scientific theories in general and their irreplaceability for the purposes which they serve.

3

The preceding remarks about the interrelations between commonsense, scientific theories and philosophical theories stand in need of elaboration; but their relevance to some recalcitrant problems in the philosophy of science can be clearly indicated. Thus, what has been said about the relation between commonsense and scientific theories throws light upon the nature of the fit between theories and experience, especially upon the notions of correspondence rules and of ' theory-laden ' concepts. Again, what has been said about the relation between philosophical and scientific theories promotes a better understanding of scientific explanation.

The restrictions imposed by classical logic and its extensions on the one hand and by specific applications of the principle of neglecting the negligible on the other upon the concepts and statements of a theory make it necessary to distinguish between them and their empirical counterparts, if any. The elementary logic underlying empirical predicates and statements is less rigid than the classical elementary logic in allowing not only exact predicates, but also inexact ones, i.e. predicates which have not only positive and negative substitution-instances, but also neutral substitution-instances which can be with equal correctness classified as positive or negative ones. The wider logic is a slight modification of a weak three-valued logic constructed by Kleene[1] and reduces for exact predicates to classical elementary logic. Deductions *within* a scientific theory, be they of theorems from axioms or be they deductions of state descriptions from axioms and state descriptions, proceed in accordance with the classical two-valued logic. Schematically the first kind of deduction may be written as $a \vdash_2 t$, the second as: $b_1 \wedge a \vdash_2 b_2$, where '$a$' stands for the conjunction of the axioms of the theory, 't' for a theorem, 'b_1' and 'b_2' for two state descriptions of a theory and '$\vdash_2$' for deducibility in classical elementary logic.

The state descriptions b_1 and b_2, being expressed in the vocabulary of a theory and being simplifications of empirical or operational counterparts, say, e_1 and e_2 are not identical with or equivalent to e_1 and e_2. This logical gap between theory and experience is acknowledged by most scientists, though understandably not fully analysed by them. It is bridged by empirical statements to the effect that within certain more or less specified contexts and for certain more or less specified purposes the theoretical state descriptions can be identified with corresponding empirical propositions. The conditions of the identification include, as has been mentioned already, the assumptions of margins of error and of freedom from known and unknown disturbances. To test a theory is thus not only to apply the deductive schema: $b_1 \wedge a \vdash_2 b_2$, but also to identify b_1 and b_2 with their empirical counterparts, the identifications being, say, $e_1 \approx b_1$ and $e_2 \approx b_2$. Any imperfection of fit between theory and experience may, therefore, be due to, and removable by changes in, either the axioms of the theory or the conditions of identification. It will not do to restore the deductive schema by incorporating the identifications into it, e.g. thus: $(e_1 \approx b_1) \wedge (e_2 \approx b_2) \wedge b_1 \wedge a \vdash_2 e_2$. For e_1 and e_2 may not fit into the framework of twoz valued logic and one of the reasons for the identifications is to replace them by b_1 and b_2 which fit into the framework.

---

[1] S. C. Kleene, *Introduction to Metamathematics* (Amsterdam, 1952), §64.

Carnap in his recent remarks on the problem of ' correspondence rules '[1] is aware of this difficulty. He suggests, in order to avoid it, the construction of a language containing ' a comprehensive logic ' and capable of accommodating not only the theoretical language, but also the observation language and the correspondence rules which bridge the gap between them. The weak three-valued logic, mentioned earlier, suitably extended by non-logical terms, predicates and postulates is such a language. However, its construction and employment only shows more clearly the difference between L_2-truth, i.e. validity in classical logic and therefore deducibility *within* theories, and L_3-truth, i.e. the wider notion of validity in the weak three-valued logic. A comparison between the truth-tables of L_2 and L_3 shows, for example, immediately that the form: $\sim p \underset{3}{\rightarrow} (p \underset{3}{\rightarrow} q)$ is L_2-true, but not L_3-true, since if p, and therefore $\sim p$, is neutral and q false or neutral, the resulting proposition is neutral. The two-valued logic with its mathematical extensions is a Procrustean bed for the empirical content put into it.

Understanding the relation between the logic underlying commonsense discourse and the more rigid logic underlying scientific theories also helps one to disentangle the variety of notions which are called ‘ theory-laden concepts ’, a disentangling which should, but rarely does, precede attempts at answering the fashionable question whether or not all concepts are theory-laden. We must first of all distinguish between intratheoretical and extratheoretical predicates. Only the former belong to the terminology of a theory and are subject to the general restrictions imposed by the logico-mathematical framework on any theory embedded in it, and to the special restrictions characteristic of the theory to which they belong. These predicates are ‘theory-laden ’ in the sense that they are incorporated into a theory. They are, properly speaking, theoretical predicates. Among them it is possible to make distinctions in accordance with their distance from experience. Some theoretical predicates are, when the theory is compared with experiment or observation, identified with empirical predicates and are in this sense the nearest to experience. The theoretical predicates not so identifiable can to some extent be stratified in terms of their logical relations to each other and to predicates which are identifiable with empirical ones. This possibility amounts to a justification of the usual distinction between theoretical predicates of different levels, but it is not a justification of the assumption that the predicates of the lowest level are, as distinct from ‘are identifiable with’, empirical ones.

---

[1] *The Philosophy of Rudolf Carnap* (Library of Living Philosophers, La Salle, 1963), pp. 958 ff.

The distinction between theoretical and extratheoretical, commonsense predicates has nothing to do with the distinction between sensible or perceptual predicates, like ‘ coloured ’, ‘ hard ’, etc., and non-sensible predicates, which like ‘ substance ’, ‘ cause ’, etc., are, as Hume has emphasised, not names of perceptual characteristics, but are, as Kant insisted, nevertheless applicable to what is perceived. Since in applying predicates like ‘ substance ’ or ‘ coloured substance ’ one interprets instances of sensible predicates rather than describes them, the non-sensible predicates may be called ‘ interpretative ’. It is possible to distinguish between different levels of interpretative predicates and to do so without committing oneself to the assumption that there are absolutely uninterpretative or descriptive predicates.

From theoretical and interpretative predicates of various levels one should distinguish what I propose to call ‘ theory-incorporating’ commonsense predicates, the structure of which shows particularly clearly how scientific theories, which arise out of commonsense thinking, seep back into it, thus modifying it in turn. Consider the empirical generalisation ‘ All magnetised pieces of iron attract iron filings', as formulated by some shrewd observer of nature in Greek times, who knows nothing of the later theories of electro-magnetism, and for whom 'magnetisation' means no more than a simple physical process involving a suitable piece of magnetic iron ore and some iron filings. The universal proposition has the form: $(x)(P(x) \underset{3}{\rightarrow} Q(x))$. After the discovery of a theory of magnetism, we learn that the empirical predicates P and Q can, with great and well-known advantages, be identified in certain contexts, with the concepts of the theory, say P' and Q'. We learn in particular that, through this identification, the theory is available in reasoning which leads from substitution-instances of $P(x)$ to substitution-instances of $Q(x)$. What we have learned, can be expressed by saying that we have recognised a new feature of these predicates, namely their identifiability with predicates of the theory, that the predicates have been replaced by different ones, or that the concepts designated by the predicates have developed. By being identified (or seen to be identifiable) with theoretical predicates and thus by having the theory incorporated in them, they do not cease to be empirical and they remain free from the general and special restrictions to which their theoretical counterparts are subject.

We can thus distinguish at least three different senses of 'theory-laden' concepts, namely (*a*) theoretical concepts, which form part of a scientific theory, (*b*) empirical, interpretative concepts and (*c*) empirical theory-incorporating concepts. It is clear that concepts may be 'theory-laden' in one sense of the term, though not in another; that most empirical concepts designated by predicates in general use are interpretative, and that as science develops and knowledge of it increases, ever more empirical concepts will be theory-incorporating. The controversial thesis of the theory-ladenness of all concepts thus gives way to some rather obvious statements.

What has been said about the constraints—the reasonable and the unreasonable ones—imposed on scientific by philosophical theories, has some bearing on the question of the explanatory power of scientific theories. Two extreme answers, namely that only scientific theories are explanations, and that scientific theories cannot explain anything, are quite unacceptable to anybody who has at one time or another been satisfied by alleged scientific or non-scientific explanations and who has preferred one scientific explanation to another. Assume that two scientific theories are logically faultless and equal in predictive power. It is then still possible that one of them should be regarded—at least by one group of people—as being a better explanation than the other.[1] This means, even when the *ceteris paribus* clause is not applicable, that the ascription of explanatory power to a theory does not depend merely on the theory's being a scientific theory, but on its being a scientific theory of a certain kind, e.g. probabilistic rather than deterministic. The analysis of theory-construction, which is of course by no means wholly rule-governed, has shown it to conform to more or less explicit regulative principles, which are expressible in the form of hypothetical imperatives: If in the construction of scientific theories certain extrascientific desiderata are to be complied with, then.... These hypothetical imperatives do not commit one to following them unless one accepts the hypothetical desiderata. But if one accepts them they become *ipso facto* requirements for scientific explanations, even though their reasonableness can only be argued on general philosophical grounds.

There is, I think, no need to summarise the discussion. One point, however, deserves further emphasis. Common sense, science and philosophy are in continuous interaction. Philosophy and science arise from modifications of common sense, enlarging and modifying both it and each other. Although philosophy, especially as reflection on science and common sense, cannot be replaced by science, it is not the unchangeable core of a changing intellectual fabric. Philosophy is perennial, but there is no *philosophia perennis*.

University of Bristol

---

[1] For examples, see e.g. ' Philosophical Arguments in Physics ' in *Observation and Interpretation*, New York (1957)

Brit. J. Phil. Sci., (1966), Volume 17, Number 4, 279–296
Printed in Great Britain

REFUTATION OR COMPARISON?*

G. C. Archibald

I

My starting point is Popper's Demarcation Rule (e.g. [5], p. 41), that a statement is to be regarded as scientific if it is refutable, and his associated injunction: try to refute. My problem is due to the circumstance that much activity in economics is devoted to empirical work with hypotheses that do not satisfy Popper's falsifiability criterion: should I judge that the hypotheses, and the associated empirical work, fall short of some ideal scientific standard, or should I judge that the methodology calls for what is impossible, and that it should therefore be altered? My own judgment is that many of the irrefutable hypotheses in economics are important, that they are incurably irrefutable for good and fundamental reasons, and that the activity of comparing them with observation is useful (too practically useful to be acceptably called metaphysics), so that it is the demarcation rule that should give way. Thus I want a methodology which will 'save the activity', as a theory is designed to 'save the phenomena', but just as a theory should exclude something, so should a non-empty methodological rule, which means that there are constraints on permissible rescue methods.

I shall propose a solution here—a new Demarcation Rule—which I think at one and the same time saves the activity and excludes from science the bulk of what Popper wished to exclude. It appears more convincing, however, to show that there is a genuine problem-situation before proposing a solution, so I shall start by discussing a somewhat stylised example which I believe to be characteristic. I shall go on to consider in more general terms why the criterion of

---

* Received 1.vi.65. This paper is based on one delivered to the British Society for the Philosophy of Science on 7 December 1964. For comments then, and since, I am much indebted to Mr R. Harré, Professor W. C. Kneale, Mr A. E. Musgrave, Dr H. Post, and Dr J. O. Wisdom. I am also much indebted to Professor R. G. Lipsey, and my former colleagues in the Staff Seminar on Methodology, Measurement, and Testing in Economics at the London School of Economics. I claim sole responsibility for any errors, eccentricities, or heresies here committed.

strict, or Popperian, refutability cannot be satisfied in much important work. I shall then argue that Popper's method of dealing with statistical hypotheses is not satisfactory, and indeed not consistent with his notion of refutability and his own demarcation rule. I shall finally propose my own solution.

It may be helpful to anticipate the argument a little. Strict refutability depends on a hypothesis absolutely excluding something, that is, assigning zero probability to some events. Like Braithwaite [2], I regard all general statements as probability statements (and any test-procedure as a sampling procedure), and refutable statements as belonging to the subset which assigns probabilities of only zero or one. I think that Popper is too restrictive in admitting only the members of that subset to science, or too optimistic in believing that all hypotheses which are in some sense scientific can be got into that subset, and I shall argue that there is good reason why we cannot follow the injunction that we ought to make our theories Popper-refutable. Thus I believe that his method of dealing with probabilistic hypotheses is unsatisfactory. As I understand him, he says ' make them refutable ' (e.g. by assigning confidence limits); but I shall argue that ' refutability ' in this sense is very different from refutability in the sense of a logical relation (that of contradiction) between a universal and a singular statement : for one thing, a ' refutation ' in the probabilistic sense is reversible, whereas one in the strictly Popperian—or deterministic—sense is not. I therefore propose a demarcation rule which will admit hypotheses that deal in probabilities other than zero and one provided that they can be critically confronted with observation. This seems to me to be consistent with the general spirit of Popper's views. It will be evident that my position is very close to Braithwaite's: I shall try to indicate the correspondences as I proceed.

Thus the argument here is (i) that the subset of general statements which are strictly Popper-refutable is too limited to be identified with the class of scientific statements and (ii) that his method of dealing with the complementary set is both *ad hoc* and incompatible with his rigorous treatment of the Popperian subset. It is also possible that the Popperian subset is empty. Braithwaite has argued ([2], p. 19) that '... complete refutation is no more possible than is complete proof', on the grounds that if a false prediction follows from the conjunction of two or more higher-level hypotheses, one can only say that the conjunction is false, and therefore *one* of the hypotheses is false. It would appear that the subset could not be entirely empty, since isolated predictions would not suffer from this ambiguity, but it might well be empty of hypotheses of scientific importance. I incline to the view that this is a matter of fact: Quine [7] apparently believes that it can be settled by *a priori* argument (for a criticism of Quine's view, see Grünbaum [3], as well as Popper [5] and [6]). I shall not pursue the question: according to the view presented here, it does not matter very much.

II

1. There are many examples in economics of serious and important hypotheses that are irrefutable but which nonetheless clearly have empirical content in some sense, and may be usefully confronted with observational evidence. Because of its fame and its practical importance, I shall take as my example the Keynesian theory of the determination of the national income but, to avoid irrelevant technical complications, I have taken the liberty of simplifying it in a way which will, I hope, reveal the problems I have in mind without doing violence to the economics. A distinguished economic theorist, Kaldor, takes as *explicanda* for his theoretical work what he calls 'stylised facts'; and I am going to stylise a hypothesis, and its test, to clarify my *explicandum*.

For expository purposes, I shall ignore the following complications:

foreign trade
government
money
price changes.

I assert, however, that Keynesian theory can easily be extended to deal with these complications (except, perhaps, the last), so that we are dealing with a deliberately simplified—and stylised—*example*, not a ludicrously naïve theory. Now, if I may be granted these simplifications, I can set out the theory.

(1) National Income may be divided into two components, A and D, thus $Y = A + D$ where
A (= 'autonomous') consists of expenditures which are predetermined for a given time-period, say one year (the elements of A will be such things as investment projects which cannot be quickly altered, housing starts, etc., and, in the real world, much government expenditure); and D (= 'induced') expenditures which depend on the current (annual) level of income.

(2) D is a *stable function of income:* $D = D(Y)$, $0 < D' < 1$

(3) Since $Y = A + D$ and $D = D(Y)$ substitution evidently yields $Y = \phi(A)$ (ϕ' is the Keynesian multiplier, k) where $\phi' > 0$.

(D' and ϕ' are the first derivatives of the functions D and ϕ respectively.)

Clearly there are two crucial steps involved in relating this theory to the real world:

(*a*) Correctly identifying A and D: if items of expenditure which we thought belonged with A turn out to vary with Y, we are in trouble. Thus the theory asserts that there is a unique and *stable* taxonomy for a given time span: once we have assigned an item to one category (a matter for observation), it will stay there.

(*b*) Finding a *stable* relation $D = D(Y)$. This there appears to be, although not necessarily a very simple one. Furthermore, it has to be estimated, and evidently not without error. This poses problems to which I shall return. (In fact, the biggest component of D is personal consumption, which proves to be much influenced by previous income. So long, however, as the function has a suitable form and is stable, we can treat C (consumption) as having a predetermined element which belongs in A—predetermined by previous income levels—and a currently induced level which belongs in D.)

Suppose that we have estimated our relation $D = D(Y)$, and hence found ϕ. Now we say 'let us test the theory', or 'tell me the level of A for any year, and I will tell you Y'. (Pick a year we did not use in estimating D, if you like, say 1966.) So we have calculated Y and an observed Y, say Y_c and Y_0. Suppose that Y_c differs from Y_0 by 3 per cent. What do we say now? It is patently absurd to say that the theory is refuted—we could just as well say that, seeing that Y_c could have any value between zero and infinity, the theory is triumphantly vindicated. So we have the problem: is a 3 per cent error large or small? Neither logic nor intuition nor revelation seem to be very helpful.

2. Now, it is obviously of crucial importance to know how the error got into the prediction, and whether error is 'in the nature of the beast' or something which we could remove or limit if we were only a little better at our jobs. The second assertion in the theory was 'D is a stable function of income'. To obtain a prediction of income, however, we need to know *what* function, i.e. we require a specific functional form *and* numerical estimates of the parameters. For simplicity, let us identify D with consumption expenditure, C. We

have available past observations of consumption and income (say, annual totals). If we start with the simplest possible relationship, we shall choose a linear form, and fit by least squares[1]

$$C_t = a+bY_t+\epsilon_t$$

getting, say, $r^2 = 0{\cdot}75$. Other theoretical considerations tell us something about b, namely that we should find $0<b<1$. We may suppose that we find $b = 0{\cdot}8$, a perfectly acceptable value. What is worrying is ϵ, or, what comes to the same thing, $r^2<1$. There are three distinct and well known sources of error:

(i) the effects of omitted variables that have a systematic influence on C, and that should be explicitly taken account of;

(ii) errors of measurement;

(iii) the classical cause—the effects of innumerable independent and individually insignificant influences.

In the case of (i), we have a 'specification error'—we have not yet related C to the right variables. This sort of error is likely to show up not only in a low r^2 but in a 'badly behaved' error term and unreliable estimate of the coefficient b. If this is the case, we clearly have something to learn. We may leave aside for a moment the question of whether the simple unlagged linear relationship has been 'refuted': let us agree that it is unsatisfactory. Now theoretical considerations—or commonsense—suggest that consumers do not adjust very quickly to changes in income, so that past levels of income exert some influence over current expenditure. So we develop a theory of consumers' expenditure which yields, say,

$$C_t = a_0+a_1 Y_t+a_2Y_{t-1}+\epsilon_t$$

together with the constraints

$$a_1+a_2<1,\ a_1,\ a_2>0$$

(The theory might also say more, e.g. that the consumer's 'memory' was less important than current experience, i.e. $a_1>a_2$.) Suppose that we now obtain $R^2 = 0{\cdot}9$ or more, and that other criteria are satisfied (we have a 'well-behaved' error term, normally distributed and free of autocorrelation). We conclude that we have found a stable relationship between C and Y. We solve for Y, and proceed as before.

---

[1] This estimator will in fact suffer from what is called 'single-equation bias', which means that, because other simultaneous relationships are ignored, the estimate of the parameter b would not converge on the 'true' or population-value as the size of the sample increased. But this does not matter for present purposes, which are illustrative.

We have not, in fact, got a refutable theory. Our solution is $Y = \phi(A, \epsilon)$. *If* ϵ *is normally distributed, with tails asymptotic to the axes, no value of* C_t, *and hence* Y_t *for given* A_t, *is absolutely excluded, however outrageous.* We cannot save 'refutability' by saying, for example, 'at least the theory would be refuted if Y ever went down when A went up': this is improbable but not impossible.

3. It is clear that much depends on the 'normally distributed error term', which requires further investigation. Remember that we have eliminated—or think we have eliminated—specification error. (We may put our question in a slightly different way. Consider our hypothesis as being that we have correctly specified the consumption function as above. Is this hypothesis now refutable?) Our problems now arise only from the second and third causes of error. In practice, it may be impossible to distinguish between the two, but they are conceptually distinct. Consider first errors of measurement.

If a theory is to be refutable, it must exclude some possible observations, as opposed only to rating them as more or less improbable. This would clearly be the case if the tails were cut off the distribution of errors. Thus if we knew that the errors on our consumption function could not exceed some value, p, we should obtain $Y < \phi(A, p)$ as an absolute limit, instead of $Y = \phi(A, \epsilon)$. Thus it would appear that, in order to obtain refutability, we only required to know the limits of error in our measuring apparatus, an apparently very reasonable scientific requirement! This conclusion is not, however, quite final, as the following considerations will suggest.

(*a*) If it is a measuring instrument we are dealing with, we presumably need another instrument—of known error—to measure its errors with, and so on. (I do not wish to dwell on this point. It appears to have been much discussed in the scientific literature, and, in some cases at least, there appear to be ways of meeting it.)

(*b*) Suppose that we can assign limits, as would be the case, for example, if we had only to deal with rounding errors, which we would take to be rectangularly distributed. It would appear that only some trivial arithmetic is now required, and, indeed, this is true in the case of the hypothesis I have been discussing. But it is easy to invent hypothetical cases of which this is not true. Suppose that my hypothesis is that $y_1 = f_1(x)$ and $y_2 = f_2(x)$, both f_1 and f_2 monotonic increasing (say, linear for simplicity), and that $f_1 > f_2$.

Fig. 1 illustrates this hypothesis. Now suppose that the limits of accuracy within which y can be observed are known absolutely. Permissible values lie within the dotted lines parallel to f_1 and f_2. The hypothesis $f_1 > f_2$ is open to decisive refutation: it is only necessary that we can make observations where $x > x_0$. Suppose now that the limits of accuracy are in proportional, not absolute, terms, so that permissible observations lie between the two pairs of dashed lines about f_1 and f_2. As drawn, all is well: a refutation could occur at any value of x. If, however, f_1 and f_2 were closer together, or our proportional limit larger, b would lie to the left of a, and refutation would be impossible. We may say, however, that this problem is to be overcome by better instruments.

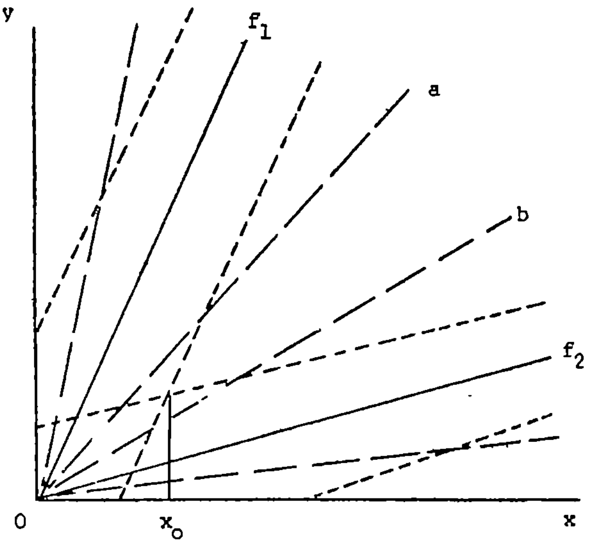


FIG. 1

(c) It is well known that, if a number of people perform the same operation of measurement, they will not all produce identical answers. Instead, there will be a distribution of results around a mean (itself an estimate of the true value). It is not obvious how we may assign absolute limits to individual error. If we could know that no one ever made a mistake larger than p, we should have the case just discussed. As it is, we dismiss 'outlying' results as 'occult', or due to personal error: we do not accept them as refutations of a theory until they are reproduced (which is to say, until the sample mean shifts).

4. Consideration of this last source of measurement error—human frailty—leads on to consideration of the third sort of error. Suppose that a variable is subject to certain important and systematic influences. It may also be subject to many individually unimportant and independent influences. There is an important statistical theorem (the Central Limit Theorem) which says that the sum of a large number of independent random variables (with finite means and variances) will be approximately normally distributed. This is why we expect ϵ to be normally distributed, and suppose that we suffer from specification error if it is not. We can argue that independence and randomness are the characteristics of free will, and that a normally distributed error term is therefore an immutable feature of social science! By 'free will' here, I mean freedom to depart from the behaviour predicted by the 'deterministic' part of the economist's equation. Consider, for example, the behaviour-equation $C_t = a_0+a_1Y_t+a_2Y_{t-1}+\epsilon_t$. Our hypothesis is not that $\epsilon_t = 0$, but that we have accounted for all the systematic influences on C_t. Each individual is free to spend more or less, in any year, for any reason: whim, marriage, parenthood, conversion to a new religion, or what you will. What we require is that there be many individuals and that these decisions be independent of one another (between individuals, at least). If they are not, then the deterministic part of the equation is wrong. (Thus suppose that a rapidly increasing Y leads people to anticipate further increases, and so to anticipate next year's increase in consumption, and vice versa. Individual decisions to depart from the level of consumption given by $a_0+a_1Y_t+a_2Y_{t-1}$ would no longer be independent, and we could not properly consign all discrepancies to ϵ. We should have to include the rate of change of income in the deterministic part of the equation.)

Thus we may claim that the exercise of free will, as defined, is normally distributed. Alternatively, we may regard this result as the consequence of a decision made by economists for their own convenience, to go on altering the deterministic part of a behaviour equation until the errors are normally distributed! (We might identify these as the 'realist' and 'Eddingtonian' views of the matter.) Which view we adopt does not presently matter. What does matter is that a normally distributed error term is a characteristic of social science, and so, therefore, is irrefutability. More generally, whenever we have a distribution that gives non-zero probabilities over the whole range $-\infty$ to $+\infty$ (or 0 to $+\infty$ if the variable can only take on positive values), we have irrefutability. The crucial point is that *this holds even in the* *absence of measurement error: strict refutability cannot be saved, in social science, by improved instrumentation, or more accurate knowledge of measurement error.* Thanks to what I call here ' free will ', the social scientist deals, whether he likes it or not, with a stochastic universe, and the injunction to make his hypotheses refutable cannot be obeyed.

III*

1. My own view is that, faced with this situation, economists have done the right thing, which is to adopt comparative methods (discussed in IV below), and the demarcation rule I shall propose hinges upon comparability. Before considering these matters, however, it seems advisable to explain why I think Popper's method of dealing with probabilistic hypotheses is unsatisfactory.

Popper's requirements are strict. ' A theory is to be called " empirical " or " falsifiable " if it divides the class of all possible basic statements *unambiguously* . . .' ([5], p. 86, my italics). Falsifiability is a ' methodological requirement '. Now, what I take him to say,[1] in the case of probabilistic statements (see [5], particularly pp. 204, 261-262) is that they are not strictly falsifiable, are ' completely undecidable ', etc., but that we can and should frame a rule to make them falsifiable. I argue that the rule required in each case is either *ad hoc*, or derived from contemplating the probable results of subsequent action, and cannot be assimilated with the rules required in the deterministic case.

2. The deterministic case is logically unambiguous: the hypothesis excludes some possible event, i.e. there is a contradiction between it and a basic statement. The decisions required are (*a*) to eschew conventionalist tricks, and (*b*) to accept the basic statement(s). Once (*b*) is taken, however, logic does the rest: there is no further room for

---

* I must acknowledge once again my greatest indebtedness to Mr A. E. Musgrave whose patient and pertinacious criticisms of earlier drafts of this paper have led me to the present version of this section. I am not content, however, with the conventional warning that he is not responsible for what is said: I explicitly assume entire responsibility for all errors that remain.

[1] I use this form of words advisedly. I do not find Popper's views on probabilistic hypotheses altogether easy to comprehend. As an alternative to extensive quotation and laborious exegesis, I have simply given here my own understanding of the matter, together with references to those pages which seem critical to me and which, I hope, the interested reader will judge for himself.

doubt, judgment, or decision.[1] In the case of a probabilistic hypothesis, we also have to take decisions (*a*) and (*b*), but all logic (the calculus of probability) can then do is tell us the probability of drawing the sample (described by the basic statement) from a universe with the characteristics we hypothesised. 'Refutability' requires a further *decision* (*c*): we must assign some (numerically valued) probability as a critical value. This is, of course, what statisticians do every day, but they do not delude themselves that they are exposing hypotheses to the possibility of refutation: they are making *provisional* 'acceptances' or 'rejections'. The circumstance that we *can* decide, that Braithwaite's 'K-rule of rejection' *can* be applied, does not appear sufficient to me to make a hypothesis Popper-refutable.

3. It should be clear that a decision in sense (*c*) is distinct from, and additional to, decisions (*a*) and (*b*); but the matter may be taken further. Consider, first, why we should take a (*c*)-decision at all. If we do not have to commit ourselves to any subsequent action which hinges upon the decision, we may merely say 'The probability that this sample was drawn from a world behaving according to the hypothesis is p', and stop. This, I take it, is Braithwaite's point: the meaning of a probabilistic hypothesis is attributable to the circumstance that we *can* assign critical limits to p. Is it not absurd to lay down the 'methodological requirement' that we shall, and that the selection of a number, perhaps quite arbitrarily, transforms the situation? There are in many cases good reasons for assigning particular confidence limits if we do wish to decide, but these are based upon considerations of the *probability* of making mistaken decisions (wrongly rejecting the true hypothesis; wrongly accepting the false): we can only decide what probabilities of the different possible sorts of mistakes we will accept, and the decision itself depends upon the computed probabilities of the samples given the hypothesis.

As an illustration of the way in which the decision in sense (*c*) may hang upon consequent action but does not establish refutability in the deterministic sense, suppose that we are jurors at a murder trial. Having weighed up the evidence, allowing for conflict, the probability of human error—and perhaps perjury—we come to the conclusion that the chances are 95 per cent that the accused is guilty. If the punishment for murder is execution, I suspect that we will find him innocent; if it is some lesser punishment, we may well waver; and, if he is being tried *in absentia*, having taken up residence in a country with which we have no extradition treaty, I suspect that we shall quickly find him guilty! It seems to me preposterous to insist that our procedure is scientific once we have determined, with an eye on the consequences, the probability of guilt at which we shall actually convict. What matters is that, in deciding the probability that the accused is guilty, we appeal to evidence; and what we want is a methodology which accommodates the appeal to the evidence without trying to force it into the straightjacket of refutability.

---

[1] Notice that, if the basic statement itself is only acceptable up to some confidence limits, not even the deterministic hypothesis is Popper-refutable. If this is in fact the general case, as I believe, then, as suggested in I above, but for a different reason, the subset of Popper-refutable scientific statements is operationally empty. Consider 'All swans are white: B has seen a black swan'. It is possible that B was drunk or colour-blind, that the swan was sick or smeared with oil, etc. Thus the basic statement must be given a probability rating, and then the 'refutability-status' of the deterministic hypothesis is that of the probabilistic hypothesis.

4. If a demarcation rule is worth having, we should, with its aid, be able to judge whether a hypothesis, and a set of suggested test procedures, belong to 'empirical science' without waiting for the specification of numerical confidence limits which may be arbitrary and *ad hoc*, or depend on, e.g. our moral sentiments about hanging. The question 'is this science?'—or 'has this got empirical content?' is, or should be, a prior question to the question of selecting confidence limits appropriate to the case: it should be decidable without reference to *particular* decision rules.

5. It may also be remarked that Popper's method of dealing with probabilistic hypotheses appears to readmit confirmation in an undesirable fashion. If it is proper to say that failure to satisfy an arbitrary cut-off point constitutes refutation, it would seem equally admissible to say that success constitutes 'confirmation'. If it is a 'methodological requirement' to call a provisional rejection a 'refutation', why should a provisional acceptance not be called a 'confirmation' or 'verification?' The answer in both cases is that this is mere name-calling, and solves no problems.

6. Suppose that, of the three causes of error discussed in II. 2 above, the third (which I attribute to the exercise of 'free will' in human affairs) were absent. Then the problem in any given case would be to decide between specification error (refutation), and measurement error, and Popper's demand, that we make up our minds in advance about the limits of measurement error, seems more reasonable than it does when the third cause of error is present. At least, it may be easier to accept

a slightly cavalier attitude to the distribution of measurement error than to the distribution of free will: the experimenter may have some control over the former, while he has none over the latter! But not too much should be made of this concession. In the first place, measurement errors, as we saw in II. 3 above, may not be accommodating. In the second place, basic statements may only be acceptable up to some probability (III. 2, note 1). It is for these reasons that I suspect that the class of hypotheses on which a deterministic, or pass-fail test, is possible may be operationally empty. But this is a subsidiary point. What matters, and what I wish to emphasise, is that the best that can be hoped for in social science is a normally distributed error term, and that methods of dealing with the distribution of pure measurement error that may seem acceptable, if a little rough-and-ready, are simply inappropriate. A simple and robust injunction, like " cut off their tails! ", will not do: the limits will be either arbitrary or determined, as we have seen, by considerations of consequent actions, and we cannot make the irrefutable refutable by this means.

IV

1. In some broad way, it is a characteristic of a science that it is concerned with observation, and constrained by observation: Keynesian theory, for example, clearly has ' empirical content ' in some loosely defined sense, is important in a most immediate, practical fashion, as giving a cure for economic depression, and has been the focus of an enormous amount of empirical work. I shall shortly suggest a Demarcation Rule; but it may first be helpful to see what economists do when dealing with a theory which requires empirical testing but is not refutable.

In this situation, economists in fact compare. In the case that there exists a rival theory, it and the Keynesian theory are set the same problem and we see which does better. Thus an old rival to the Keynesian theory that I have presented is simply

$$Y = cM$$

where M is the money supply, as determined by the monetary authorities, and c is a behavioural constant to be estimated from past observations of M and Y. Suppose that prediction from the equation proves to be less accurate than from the Keynesian equation (as there is reason to expect to be the case). We have no further possibilities in the way of 'crucial experiments', or critical comparisons between two rival paradigms; but we may still subject the Keynesian theory to further comparison tests by *inventing* hypotheses for the purpose. Thus the economist challenges his important hypothesis to do better than the *ad hoc* hypotheses he can readily manufacture for test purposes, the so-called naïve hypotheses. Here are some examples:

(i) $Y_t = Y_{t-1}$	(this year's income is the same as last year's, surprisingly hard to beat!)
(ii) $\Delta Y_t (= Y_t - Y_{t-1}) = \Delta Y_{t-1} (= Y_{t-1} - Y_{t-2})$	(the absolute increase of this year's income over the last year's is equal to last year's over the year before's)
(iii) Extrapolation from the linear regression $Y_t = a+bt$	(as in (ii), but we use several years' observations to estimate b, the average annual absolute change)
(iv) Extrapolation from the regression $\log Y_t = \alpha+\beta \log t$	(as in (iii), estimating the annual rate of change in place of the absolute change),

and so on. Characteristically, we shall beat some of the naïve hypotheses quite easily, but not all. To beat all those listed here with a Keynesian model, however, we only need autonomous expenditure to fall: barring some 'bad luck' with the errors, income will fall and the Keynesian equation will predict that fall, whereas none of the naïve hypotheses can predict a change in direction. Even so, we could still, if we wished, construct a naïve hypothesis that would win: we can always find some polynomial with turning points in the right places. If we do this, however, we allow the naïve hypothesis fewer degrees of freedom than its serious rival, which makes the comparison 'unfair' and pointless, for if we do not adopt the rule that the naïve must have no fewer degrees of freedom than its serious rival, the former can always be selected to win.[1] Thus we have at least one handy rule of thumb for comparison testing: and we could obviously manufacture more, for example, by assigning different weights to discrepancies of different magnitude but the same sign to those of different sign.

2. Now, in fact, the sorts of comparisons just discussed do not exhaust the matter, nor are they unusual, isolated, or freakish: they

---

[1] On this point I am much indebted to the members of the Society who participated in the discussion.

actually belong to a large and familiar family. Textbooks on statistics usually have chapters called ' Hypothesis Testing ' which actually turn out to be about the comparison of hypotheses. The hypothesis to be tested is irrefutable on sample evidence: it is tested against a rival, which may be a genuine rival or merely the artifact ' not H_0 ', and a conclusion is drawn about the probability of drawing the sample of observations at hand from a population described by H_0 rather than its rival. In statistics, which deals with samples, and therefore *all* testing of universal hypotheses, the comparison of irrefutable hypotheses has been greatly refined and formalised. (Notice that our familiar squared correlation coefficient, r^2, is a comparative measure: it tells us how much better our fitted relationship predicts the independent variable than does its own mean.)

V

1. There is in fact nothing particularly novel about the notion of comparison. Implicitly or explicitly, it is central to statistics. And Popper has a good deal to say about the comparison of hypotheses as a useful activity (although I do not think that he discusses the deliberate invention of *ad hoc* hypotheses to facilitate comparison). Now, I suggest that the ideal-type, or Popperian test, the decisive pass-or-fail test, is a particular case of comparison in which we not only test a hypothesis against a rival (here, its negation), but think that it is possible to obtain a final decision (final, of course, only if the decision is adverse). Now, refutability depends upon a hypothesis absolutely excluding some logically possible event or events. Thus ' All swans are white ' is Popper-refutable because we have a degenerate probability distribution with mean at colour = white and zero variance; trivially, that is, the tails of the distribution cut the axis at the mean value which has probability one, all other values having probability zero. My notion, like Braithwaite's, is that this is a limiting case of *comparability*, the case in which a decisive rejection is possible (in principle, that is: there is still the problem that we may only be able to establish the probability of a supposed counter-example). From the point of view of a social scientist it is an ideal type, or perhaps, given free will, not even that.

Now we require a demarcation rule which admits as scientific a statement which can be compared, not necessarily decisively, with others by appeal to observation, and which will exclude it if comparison can only be done by appeal to taste, faith, or intuition. Notice that comparison ' by appeal to observation ' satisfies the canon of being intersubjectively testable.

I suggest the following Demarcation Rule, *that we call a statement—or hypothesis—scientific if we may, at least in principle, compare its probable truth or falsity with that of another statement by appeal to observation (reference to facts)*. As an alternative formulation with the same intent, we might use Popper's rules ([6], pp. 232–234) for comparing the truth or falsity content of two theories, and say that a theory is scientific if it is possible to make a (probabilistic) comparison of its truth or falsity content with that of another theory. The case in which the probabilities are limited to the values zero and one is obviously included as a limiting case. This demarcation rule has some fairly serious (non-Popperian) implications which require discussion and defence. A few points, however, may be quickly noticed. First, the general intention is to put on one side of the demarcation line all statements that have empirical content, where ' having empirical content ' ≡ being potentially comparable. Second, for the impossible injunction to make hypotheses refutable, we may substitute ' make them comparable '; and, similarly, we may substitute ' compare! ' for ' try to refute! '

2. The proposed demarcation rule admits as scientific existence statements such as ' There are sea-serpents ', since ' there are ' can obviously be compared with ' there are not '. Now, this is a fairly serious matter since it abolishes what has been called (Popper, [6], p. 41, n. 8) a ' fundamental asymmetry '. ' There are ' is, in principle, confirmable (one specimen) but not refutable; ' there are not ' is refutable (one specimen) but not confirmable. On the refutability rule, ' there are not ' is scientific, ' there are ' metaphysical. I confess that I find the asymmetry more a matter of grammar than a fundamental matter of science and non-science.[1] That the two sentences have an asymmetrical relation to verification and refutation is clear enough; but that this should have ' fundamental ' status seems a little exaggerated. Suppose that you, the reader, and I differ on a matter of fact, the existence of sea-serpents: are the rules of debate such that one of us must be making a non-scientific assertion? This might seem absurd, and might also have a serious practical consequence: since one of us would win by a walk-over, we should have no incentive to put the matter to the test of enquiry. One cannot, after all, conduct an experiment to test a metaphysical assertion! It seems to me that we

---

[1] I must also confess that I used to accept Popper's doctrine of sea-serpents quite uncritically: see Archibald [1].

should not only fail to look for sea-serpents, but for certain stars and elements too. Perhaps, as a believer in sea-serpents, I can rescue myself by saying 'I doubt your hypothesis that there are no sea-serpents: please test it'; but I may not say 'there are'. It does not seem very useful to require that a statement with obvious empirical intent be given this sort of round-about expression before it is respectable.

It is, in fact, easy to imagine a biologist saying 'There ought to be sea-serpents. Creatures which we call serpents crawled out of the primeval ooze and adapted, some to dry land and some to fresh water. Some should have settled for salt water. If not sea-serpents, sea-serpent fossils.' (If he found them, there would doubtless be a hot dispute over the crucial characteristics of serpent-ness. But that is another story, about 'conventional decisions' always being necessary to establish refutations or comparisons.) According to my rule, 'there are sea-serpents' is a comparable statement. If we look hard for them, without success, then 'there are no sea-serpents' becomes what Popper would call a well-corroborated hypothesis. I would only add: 'corroborated by comparison with its rival (which, in this instance, is simply its negation)'. As a second example, let us consider Gods and devils. Popper's rule admits 'There is no God' *if, and only if,* there is some conceivable experiment which could show that there is one. So does mine, which consequently admits, *on the same terms,* 'There is a God'. The effect of my rule is this: *if a statement is admissible, because we can appeal to observation to judge its (probable) truth, then so also is the statement which would replace it, whether decisively or tentatively, as a closer approximation to the truth if the evidence were adverse.* If I felt Pythagorean, I should call this rule the 'fundamental symmetry!' It follows, of course, that I admit statements about both the existence and the non-existence of the devil if and only if some experiment is proposed. Popper, in his discussion of 'there is a devil' ([6], p. 249), considers a patently bogus experiment which we know will fail to raise the devil. He concludes that the trouble lies with the existence form of statement. But if anyone who believes in devils will stake his belief on the outcome of a feasible experiment, surely we should take him at his word instead of telling him 'until you change your grammar, you are talking metaphysics in spite of your willingness to go to the court of observation'.

3. The Popperian position appears to be that induction is impossible and/or wicked when done. The word has been used in more than one sense, but in one sense at least, 'induction' appears to be simply a

method of comparison. A scientist who is comparing the empirical success of his hypothesis with that of another—whether a serious and pre-existent rival, the null hypothesis, or an *ad hoc* artifact—by any of the methods modern statistics have put at our disposal may describe his proceedings in inductivist language, and claim that the successful theory is 'inductively confirmed' or 'verified'. We may ask him to moderate his language, but need not ask him to proceed otherwise. The 'problem' posed by the use of an 'impossible' method having apparently yielded great results (Bertrand Russell called it a scandal!) disappears if we become 'comparativists'.

4. It remains to point out that comparison is multi-dimensional. In the first place, we compare rival theories by reference to such criteria as scope, generality, elegance, etc.: we ask, e.g. if one is 'about more things' than another, or carries excess baggage in the form of unnecessary elements, or connects more satisfactorily with other parts of our theoretical structures.[1] We should not be surprised if a theory which is superior on some counts is inferior on others! In the second place, when we compare theories with observation, we commonly find more than one criterion. Thus we may ask which better accounts for, e.g. total variance, or for turning points, or for amplitude of fluctuations. Once again, we should not be surprised if the theory which does better by some criteria does worse by others. The case of vector-dominance is the rare and lucky one in which one theory at last wins all down the line, so that we may reject its rival without waiting for the refutation that never occurs.

It will thus be seen that comparisons are frequently indecisive, not merely because each comparison is itself indecisive, being probabilistic rather than deterministic, but because the complexity of theory and observation give rise to multiple criteria for comparison. Thus when we say that one theory is 'doing better' than a rival, we refer specifically to that set of (individually indecisive) comparisons that has already been carried out. It is this complexity of comparison, rather than pigheadedness, that accounts for the well-known circumstances that we frequently disagree over theories even after a good deal of relevant empirical work has been done! It accordingly seems that Kuhn's view

---

[1] On these grounds, the naïve hypotheses of IV. 1 above are clearly inferior to their Keynesian rival. Evidently comparison between two well-articulated theories is more interesting than comparison between a theory and an *ad hoc* artifact; but the artifact can be useful, and its utility serves to illustrate further the variety of occasions for and means of effecting comparisons.

Brit. J. Phil. Sci., (1966), Volume 17, Number 4, 297–311
Printed in Great Britain

ON MODELS*

J. W. SWANSON

I

THERE is a voluminous literature on models, but it all really reduces down to a phrase: ' imputation of structure '. A good deal of what has been written on the subject of models, one feels, has really been directed at saving this concept from triviality. For if the use of models comes down to saying merely that two areas of discourse share a common structure—are ' isomorphic to one another ' is the usual way of putting it—then one has already said essentially all that can be said about models, and what remains must be explained in terms of such undefined and rather obscure notions as ' root metaphor ' (Stephen Pepper[1]) and ' open texture ' (Mary Hesse[2]), or else in terms of something like ' heuristic value '. But since neither of these sorts of explication is philosophically satisfactory—the one because it is metaphorical and vague, the other because it is psychology—the philosophy of model theory must fall back on the relatively sterile notion of shared or isomorphic structures. Intuitively, however, one feels that there must be something wrong with this position. For on such an account, once the imputation of structure has been made, the isomorphism grasped and made formal, the logical consequences traced out, the model must cease to be pregnant, cease to be a source of insight. It becomes sterile, a dessicated fossil of interest only to the paleontologist of theory formation.

This will not do. It we are correct in thinking that the role of models is absolutely crucial in science, that they are of more than ' psychological ' interest to the philosopher of science, then a more adequate theory of the role of models is required. In this paper we shall attempt to show where contemporary model theory has gone astray, and to supply the conceptual machinery necessary to an adequate theory of models.

---

* Received 23.ii.66

[1] Stephen C. Pepper, *World Hypotheses* (University of California Press, 1942)

[2] Mary B. Hesse, *Forces and Fields* (Thomas Nelson and Sons, 1961), pp. 8 and 23; following F. Waismann in ' Verifiability ', in A. Flew (ed.), *Logic and Language, First Series* (Basil Blackwell, 1952), p. 119

2

The use of models in science is generally acknowledged to involve the following. An original theory or field of investigation T_1 has been delimited and probed in preliminary fashion, but remains incomplete and contains gaps. Subsequently, a second more developed theory T_2, which we shall call ' the model ', is found to share certain structural similarities with T_1, which we shall call either ' the interpretation ', or ' the theory '. More specifically, let us suppose T_1 to refer to a finite or denumerable set of elements, or relata (denoted by the variables ' x_1 ', ' x_2 ', . . ., ' x_n ', . . .) ordered n-tuples of which will constitute the relations denoted by ' R_1 ', ' R_2 ', . . ., ' R_j ' of T_1. Similarly, suppose T_2 to refer to elements denoted by the variables ' x'_1 ', ' x''_2', . . ., ' x'_m ' . . ., and to relations denoted by ' R'_1', ' R'_2 ', . . ., ' R'_k ' ($k > j$). Let us suppose further that it is possible to find a function g expressing the ' rules of correlation ' (Max Black[1]) which associate the elements of T_1 with those of T_2 and such that $g(x_i)=x'_i$ and $g^{-1}(x'_i)=x_i$. The use of T_2 as a model for theory T_1 then occurs in the following manner. Some among $R_1, R_2, \ldots, R_j$ are observed to have analogues among $R'_1, R'_2, \ldots, R_k$, in the sense that certain instances of the latter are found to occur if and only if certain instances of the former occur. The structure of the better known relations of the model T_2 are then imputed to the theory T_1 by the inverse g^{-1} of the mapping function g. If this extrapolation holds, i.e. if one can experimentally confirm that when $R'_i x'_{i_1} x'_{i_2} \ldots x'_{i_j}$ holds in T_2, then $R_i x_{i_1} x_{i_2} \ldots x_{i_j}$ holds in T_1, then T_2 is said to be a genuine, in the sense of pragmatically warranted, model of T_1. For example, it might be the case that for certain elements x_1, x_2, x_3 of T_1 it has been found, say, that $R_1 x_1 x_2$ and $R_1 x_2 x_3$ hold. Suppose further that their analogues $R'_1 x'_1 x'_2$ and $R'_1 x'_2 x'_3$ are found to hold in T_2. Finally, let us suppose that in the more fully developed model T_2, it is known that R'_1 is transitive, i.e. that $R'_1 x'_1 x'_2$ and $R'_1 x'_2 x'_3$ jointly imply $R'_1 x'_1 x'_3$. Extrapolation of this structure to the analogue relation R_1 in T_1 yields that $R_1 x_1 x_2$ and $R_1 x_2 x_3$ jointly imply $R_1 x_1 x_3$—in other words reveals the hitherto undiscovered transitivity of R_1. If the structure of all of the relations of the model T_2 carry over in this fashion to the interpretation T_1, then one says that there exists an isomorphism of structure between model and interpretation.

---

[1] Max Black, *Models and Metaphors: Studies in Language and Philosophy* (Cornell University Press, 1962), chapters 3 and 13. Black's work on the notions of both model and metaphor is the best that has been done. The stimulation for a number of the ideas in this paper are due to his work.

This is the classical characterisation of models. It can be put into non-technical language as follows. Certain pairs of theories are guessed to have a like structure, to be isomorphic to one another. One then imputes the structure of the better known theory to the less well known, or more primitive theory. If the imputation 'takes' —i.e. if one can extrapolate as in the example above from properties of relations in the more developed theory to properties of their analogues in the less developed one—then the better developed theory is said to be a model for the less developed theory. Or even more succinctly: one suspects, on the basis of certain inconclusive clues, an isomorphism of structure between a rudimentary and a finished theory; one then imputes the structure of the latter to the former. If the suspicion was correct, one immediately gets (via the inverse mapping) analogues in the rudimentary theory of all the known deductions in the finished theory. The formula is, roughly: 'suspicion of isomorphism, imputation of structure'. We can extract this little formula from almost any account of the function of models in scientific theory formation. From the *locus classicus* for the theory of models—Braithwaite's *Scientific Explanation*[1]—we get the following picture.

> If we have before us two deductive systems which are each interpretations of the same calculus, in the first of which the interpretation of the initial formulae containing the theoretical terms is epistemologically prior to that of the derived formulae not containing these theoretical terms, whereas in the second interpretation the reverse is the case, the derived formulae being the epistemologically prior, the first deductive system will be said to be related to the second deductive system as *model* is to *theory*. . . . A theory and a model for it, or a model and a theory for which it is a model, have the same formal structure, since the theory and model are represented by the same calculus.

The rather obscure notion of 'epistemological priority' aside, this theme of isomorphism of structure emphasised by Braithwaite is just about ubiquitous in the literature of models. Thus we find Hesse[2] saying that the 'most obvious property of a satisfactory model is that it exhibits an analogy with the phenomena to be explained, that is, that there is some identity of structure between the model and the

---

[1] R. B. Braithwaite, *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science* (Cambridge University Press, 1953), pp. 89-90

[2] Op. cit. p. 22

CALCUTTA CENTRAL LIBRARY

phenomena'. And Black[1] '... the key to understanding the entire transaction [i.e. use of a model] is the identity of structure that in favourable cases permits assertions made about the secondary domain to yield insight into the original field of interest'. The crux of these separate accounts is contained in the key phrase, 'identity of structure'. In what follows we shall argue that this expression is misleading in the sense that it obliterates the essential asymmetry of model and theory—an asymmetry which is the chief source of insight in the use of models.

3

Without exception, to our knowledge, all of the current accounts of models in terms of isomorphism of structure between model and theory have construed the relation of isomorphism as symmetrical: if A is isomorphic to B, then B is isomorphic to A. Indeed, symmetry of isomorphism is often considered so obvious as to merit only a sidelong glance, or a parenthetical remark. Thus, May Brodbeck[2] merely notes in passing: 'The notion of "model" as isomorphism is obviously symmetrical.' The matter goes even more smoothly for expositions of models where the notion of 'identity of structure' is made explicit; they are excused even the burden of a parenthetical remark, since identity is known to be symmetrical. Braithwaite,[3] for example, simply notes that a 'theory and a model for it ... have the same formal structure, since theory and model are each represented by the same calculus'. Moreover, the standard mathematical definition of isomorphism as a one-one correspondence is obviously symmetrical. The mathematical definition runs as follows. Two theories T_1 and T_2 are said to be isomorphic if there exists a one-one function g mapping the structure (i.e. the relations and functions) of T_1 onto T_2 such that the inverse g^{-1} of the function maps the structure of T_2 onto T_1.[4] Here again, whatever x and y you pick, Ixy if and only if Iyx (using 'I' for the dyadic relation of isomorphism). It would seem scarcely reasonable, then, to question the assumption of symmetry for isomorphism as that relation has been used in the theory of models.

---

[1] Op. cit. pp. 230-231

[2] May Brodbeck, 'Models, Meanings, and Theories'. In Llewellyn Gross (ed.), *Symposium on Sociological Theory* (Harper and Row, 1959), p. 379

[3] Op. cit. p. 90

[4] Cf. Solomon Feferman, *The Number Systems: Foundations of Algebra and Analysis* (Addison-Wesley, 1964), p. 56; and Elliot Mendelson, *Introduction to Mathematical Logic* (D. Van Nostrand, 1964), p. 90

But *there is* a definition of isomorphism so framed that the relation is not symmetrical—i.e. such that it is not the case that '*Ixy*' implies '*Iyx*'. We are referring to Nelson Goodman's characterisation of extensional isomorphism, in *The Structure of Appearance*,[1] in terms of replaceability of ultimate factors. And it is our contention—one of the two main points of this paper—that Goodman's definition, the non-symmetrical one, is the definition of isomorphism relevant to model theory. Let us see why this is the case.

Goodman, following standard usage in logic, defines relations as ordered n-tuples of elements. (Actually, this is not his terminology, but the difference is not essential.) The most important aspect of Goodman's characterisation of relations, for our purposes, is that the elements filling in the *n* slots, say, of an *n*-adic relation, *need not* be simple elements (or 'ultimate factors', as Goodman calls them), although they *may be*. For example, $\langle a, b\rangle$ and $\langle\langle a, b\rangle, b\rangle$ are both perfectly good instances of dyadic relations, although the second one has a complex element, itself an ordered pair in fact, as first element. On this account, then, the extensions of dyadic relations (say) are in general sets of ordered couples the members of which are decomposable into further ordered *n*-tuples, these being decomposable into other *n*-tuples, and so on. But the process terminates—i.e. decomposition cannot proceed indefinitely. Now having established this, Goodman goes on to define isomorphism of relations as follows:

> A relation R is isomorphic to a relation S in the sense here intended if and only if R can be obtained by consistently replacing the ultimate factors in S.[2]

Suppose S consists of the ordered 1-tuples, K, L, M, N, while R consists of the following ordered 2-tuples (i.e. ordered pairs): $\langle a, 1\rangle$, $\langle a, 2\rangle$, $\langle b, 1\rangle$, $\langle b, 2\rangle$. Now R can be construed as having been got by replacing the ultimate factor K of S with $\langle a, 1\rangle$ of R, the ultimate factor L of S with $\langle a, 2\rangle$ of R, and so on. So R is isomorphic to S. But as Goodman's definition clearly specifies, it does not follow that S is isomorphic to R.

> Since the replacing elements need not be ultimate factors, this sort of isomorphism is not symmetrical; for if R is isomorphic to S, still

---

[1] Nelson Goodman, *The Structure of Appearance* (Harvard University Press, 1951). Cf. in particular chapter 1, sections 3 and 4

[2] Ibid. p. 12. Actually, Goodman uses unordered pairs '*a* : 1', '*a* : 2', etc., at this point. In the present context the distinction is inconsequential, although we shall need it in section 6.

> there may be no way of replacing the ultimate factors in R so as to obtain S.[1]

Goodman is concerned with theory construction, not with uninterpreted calculi. It is for this reason that he frames his definition of isomorphism in such a fashion that the possible *sub-structure* of the relations in a given theory are not lost sight of in what is, for his purposes, the excessive abstraction of the mathematical characterisation of isomorphism. Goodman is of course aware of the classical definition of isomorphism—the sense in which the sub-structure of R is abstracted from and allowed to drop out—as he indicates in the sentence immediately following the passage just quoted. ' Nevertheless if R can be obtained by consistently replacing the ultimate factors in S by certain elements it will also be true that S can be obtained by replacing those elements in R [e.g. the complex ones ⟨*a*, 1⟩, ⟨*a*, 2⟩, etc., above] by the correlated ultimate factors '.[2] This last sentence expresses the crux of the mathematical definition of isomorphism.

In terms of model theory the significance of Goodman's definition is just this. We can say that T_2 (the model) is isomorphic to T_1 (the interpretation, or theory), but we cannot say that T_1 is isomorphic to T_2. In effect, this is to construe the relata x'_1, x'_2, . . . etc. of T_2 (see above, section 2) as complex, or non-ultimate factors. Now this point is absolutely crucial to an understanding of the logic of models. For only if we abstract from the excess meaning, only if we discount the richer substructure of the model, can we say with Braithwaite, Brodbeck, Nagel[3] and the rest that the employment of models always involves just one calculus with two theories. But as was remarked above, such a position regarding the logic of model theory is uncontroversial, unfruitful, and generally of little interest. What has always been at the bottom of the problem of models, and what has inevitably caused the most perplexity for model theorists, has been the problem of how to avoid having to accept the counter-intuitive position that models are conceptually vacuous devices, ' substitutes for . . . taking the scientic theory straight . . . props for feeble minds . . . surrogates for some precedure '.[4] For once the calculus C is known, tracing out its consequences in T_1 and T_2 is a routine job, identical for both theories in everything except nomenclature. No, the notion of a model is non-trivial *only if we impute deeper structure to the model than to its interpretation.*

---

[1] Nelson Goodman, *The Structure of Appearance* (Harvard University Press, 1951), p. 13

[2] Ibid. p. 13

[3] Ernest Nagel, *The Structure of Science: Problems in the Logic of Scientific Explanation* (Harcourt, Brace and World, 1961). Cf. especially pp. 95-97

[4] Black, pp. 235-236

It must not be inferred that this has gone entirely unnoticed in the literature. A number of philosophers have admitted the need for assuming greater complexity for the model than for its interpretation. But strangely enough, such claims have invariably been coupled with a ' one calculus two theories position ' regarding the sort of isomorphism that must obtain between model and interpretation. Such a view of course, implying identity of structure, obscures the fundamental point that is brought out in Goodman's non-symmetrical definition of isomorphism. The reason, incidentally, for the wide-spread and almost totally uncritical acceptance of the symmetrical isomorphism position is not difficult to pin-point. Almost all contemporary writers on models—Nagel, Brodbeck, Achinstein, Braithwaite (but perhaps not Black)—assume, as point of departure for their analyses, the hypothetico-deductive position on the nature of scientific theories. That is, they assume in their discussions that the sciences to which their theories of models apply have been regimented into axiomatic form as universally quantified formulae of the first order predicate calculus, and that these formulae express general law-like connections, such that when conjoined with other formulae expressing certain initial conditions, they permit the deduction of specific observationally confirmable sentences. To invoke the limp old paradigm, if our science contains as general principle the universal sentence '$(x)(Cx \supset Bx)$' (all crows are black), and we adduce as initial condition 'Ca'—i.e. the thing a is a crow—then we can deduce 'Ba' (a is black). But of course science in general is not axiomatised, give or take a portion here and there organised by a Woodger or a Carnap. So when speaking of imputing the structure of some theory T_2 to some other theory T_1, the model theorist must mean ' the structure *if axiomatised* ' of T_1 and T_2. Let T_2 be a classical theory of planetary motion, and T_1 be a theory of sub-atomic structure. To impute the structure of T_2 to T_1, as was done in the Bohr-Rutherford theory of the atom, is to suppose that an *axiomatised* T_2 exhibits a structure, which, once imputed to T_1, will generate universal principles in T_1 that will allow, when conjoined with certain (experimental) initial conditions, the deduction of observationally confirmable conclusions. As we know, this particular imputation did ' take ', leading to one of the great forward strides in science. Our point is that if T_2 had been fully axiomatised, it would have contained more structure—the un-imputed sub-structure referred to above—than T_1. One could have got T_2 by replacement of 'ultimate factors' in T_1, but not the other way round. The isomorphism would not have been symmetrical.[1]

In a recent and perceptive article,[2] Peter Achinstein has come very close to realising the point we are attempting to make. Criticising what he characterises as the 'identical calculi' view (in our terminology the 'one calculus two theories' view), he remarks that

> it precludes by far the largest number of cases in which an object described in a particular theory is compared with another object *only in certain* respects. Thus, when an analogy is drawn between the Bohr atom and a planetary system, only certain properties of atoms and solar systems are relevant in the description of the analogy [i.e. model] (e.g. orbits with relatively large masses at centre). Other properties attributed to atoms by the Bohr theory (e.g. quantum jumps with electromagnetic radiation resulting) are in no way considered part of this analogy. ... Numerous characteristics of planets (e.g. their reflection of light and differences in diameter) do not correspond to properties ascribed to electrons by the Bohr theory.[3]

Achinstein then suggests the following reformulation of the identical calculi position.

> The set of propositions describing an analogue [model] if axiomatised, will not in general have the same calculus as the corresponding theory [interpretation]. At best it can be expected that some propositions in the theory will be formally similar to certain laws governing the analogue.[4]

Unfortunately, Achinstein's reformulation is not entirely adequate. For what he says concerning isomorphism of model and theory can be summarised as follows. Given two theories T_1 and T_2, let C_1 and C_2 be respectively the sets of their axioms, $C_1 \neq C_2$. Then there exist sub-sets C_1' and C_2' of C_1 and C_2, these sub-sets themselves constituting theories that *are classically isomorphic* to one another, i.e. isomorphic in the symmetrical sense. So the identical calculus view is, in effect, reinstated. We must look elsewhere, then, for an explication of the non-symmetrical view that will make intelligible a theory providing an autonomous and non-eliminable role for models.

---

[1] If this sentence still sounds strange, perhaps contradictory, recall that 'isomorphism' is being used as a technical term, in Goodman's sense.

[2] Peter Achinstein, 'Models, Analogies, and Theories', *Philosophy of Science*, vol. 31, no. 4 (October, 1964), pp. 328–350

[3] Ibid. p. 336

[4] Ibid. p. 337

4

Some philosophers have suggested assimilating the notion of model to that of metaphor, in order to account for the unique and non-eliminable role that models play in theory formation. Among interpretations in this vein, that of Max Black is far and away the clearest.

> Certainly there is some similarity between the use of a model and the use of metaphor.... And the crucial question about the autonomy of the method of models is paralleled by the ancient dispute about the translatability of metaphors. Those who see a model as a mere crutch are like those who consider metaphor a mere decoration or ornament. But there are powerful and irreplaceable uses of metaphor....
>
> A memorable metaphor has the power to bring two separate domains into cognitive and emotional relation by using language directly appropriate to the one as a lens for seeing the other; the implications, suggestions, and supporting values entwined with the literal use of the metaphorical expression enable us to see a new subject matter in a new way.[2]

Elsewhere, in another penetrating analysis, Black offers powerful arguments for his ' lens ' view of metaphor. It should not diminish the importance of Black's insightful theory of metaphor—the most valuable contribution, in our view, to the recent literature of this topic—to point out that the burden of his analysis of metaphor is itself conducted in metaphorical language. For on Black's view, a good metaphor acts much like a filter, or lens. To say that man is a wolf is to see man in terms of our commonplace knowledge about wolves—to see that man, too, can be cruel, rapacious, treacherous, etc. It is to realign, readjust, and focus anew our commonplace knowledge about man, in terms of how he appears through the filter, the lens, the conceptually tinted spectacles of what we know about wolves. This is a capsule picture of course, no doubt something of a caricature, of Black's position on metaphor. But it will do for our purposes, since we want only to argue that this view of metaphor furnishes the point of departure from which an adequate theory of the logic of models can be constructed. What is lacking is a way to transform Black's metaphorical account of model and metaphor into something approaching an empirical and testable theory. Although we shall not present a finished and detailed theory of such a transformation of the metaphorical content of models, we shall sketch out in broad strokes the outlines of a theory susceptible to further refinement and testing. The missing ingredient needed to extend Black's analysis to an empirical basis, we shall argue presently, is the notion of a rational preference ranking.

---

[2] Op. cit. pp. 236-237

5

Black's theory of metaphor runs as follows. The typical metaphor —let us use Black's own example, ' Man is a wolf'—is constituted of a principal subject (' man ') and a subsidiary subject (' wolf'), with each of which are connected in the mind of the ordinary speaker (of English in this case), a certain system of ' associated commonplaces '— i.e. a set of readily evoked ' current platitudes ' (not necessarily all true) concerning men and wolves. On this account both lupine expert and man in the street, in so far as they are native users of English, share a common pool of platitudes concerning wolves: they are treacherous, carnivorous, cruel, savage, etc. And the same for ' man '. To say that man is a wolf, then, amounts to looking at man through the wolf-lens, i.e. in terms of the associated commonplaces connected with ' wolf'. ' The metaphor selects, emphasises, suppresses, and organises features of the principal subject by implying statements about it that normally apply to the subsidiary subject '.[1] To see man as wolf is to recognise that men are treacherous as well; it is to glimpse that a particular associated commonplace of the principal subject, the commonplace expressed by the sentence ' man is rational ', need not always be true. Under the influence of the metaphor this particular platitude recedes into the background, as it were, while others—' man is capable of treachery ', ' man preys on man ', etc.—are drawn into the foreground. This is the filter view of metaphor. As remarked above, it explicates ' metaphor ' via metaphor. Can we improve on this?

Well, what does a filter do? It allows some thing to pass, disallows others. It organises the things which pass through it in some sort of systematic fashion. It selects some things over others. It *rearranges* things. What a filter does is not so much to exclude things, although it does that too, as to let them pass in a certain order so that those preferentially chosen will come first, those least preferred will come last, and the rest will be ranked in order in between. (Think of a flour sifter, allowing the finest grains to pass first, then the larger ones, retaining only the lumps.)

---

[1] Op. cit. pp. 44-45

It is clear that the characterisation we have just given of a filter would serve equally well as an intuitive account of a rational preference ranking. We can characterise a rational preference ranking (subsequently, ' RPR ') as follows. Suppose an experimenter E interrogates or observes a community of language users *x*, *y*, *z*, . . . etc., with regard to their preferences concerning certain classes of sentences in their language. E might directly question a language user *x* as to his preference with regard to two sentences *a* and *b*, or he might simply observe *x*'s behaviour in the appropriate contexts, and from this infer *x*'s preference. But note that it is a pair of linguistic entities, not extra-linguistic objects, that we are supposing E to confront *x* with. Thus our account follows that of R. M. Martin[1] in construing preference as holding between sentences, rather than between material things, qualities, pleasures and pains, and so forth. For our purposes, a complete behaviouristic account of ' preference ' is not necessary, although of course for the hypothetical experimenter E it certainly would be. We have a number of alternatives for interpretation of ' preference '. ' Suppose E presents the subject *x* with two sentences *a* and *b* of the language L at time *t*. And suppose he asks *x* which of the sentences *x* regards as the *more probable*, or in which he *more firmly believes*, or whether he is willing to accept one more than the others, or whether he is more willing to *base his actions* on one than on the others.'[2] Our own choice of preference relation (one for which powerful arguments can be adduced) would be a negatively framed one—' *x* prefers *a* to *b* at *t* ' means ' *x* is less willing to relinquish *a* from his corpus of knowledge than *b* at time *t* '. This particular interpretation of preference would be the one most appropriate, for example, to a Duhemian theory of the acquisition and subsequent revision of scientific lore. But this is a matter we need not argue here.

Now to say that the individuals interrogated by E display RPR's with respect to a given set of sentences is to say that each individual *x* would, if presented by E with two sentences *a* and *b*, either prefer *a* to *b*, or *b* to *a*, or else be indifferent with respect to *a* and *b*. That is, he must choose exactly one of these alternatives. (We need not introduce a separate primitive predicate for ' indifference ', construable intuitively as ' ranking at the same level ', since it can be defined in terms of preference—as ' $\sim P\,x\,a\,b\,t \wedge \sim P\,x\,b\,a\,t$ '—taking ' P ' as our tetradic preference predicate.) Moreover, if x's choices are to constitute a rational preference ranking, then preference must be transitive—' $P\,x\,a\,b\,t \wedge P\,x\,b\,c\,t. \supset P\,x\,a\,c\,t$ ' holds—and asymmetrical —' $P\,x\,a\,b\,t \supset \sim P\,x\,b\,a\,t$ ' holds—while indifference must be both transitive and symmetrical. Given that these conditions hold for x's ranking of a certain set K of sentences at time t, we say that x exhibits a RPR with respect to K at t. Intuitively, this is to say that for every sentence a of K at t, x ranks a relative to every other sentence b of K, with a either above, below, or at the same level as b. Clearly, the notion of a rational preference ranking is no more than a crisp and formal explication of the commonplace notion of ' grading '.

---

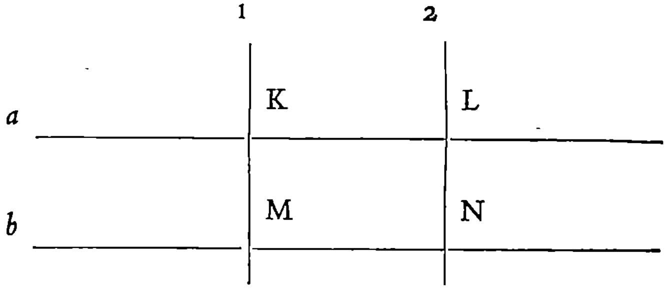
[1] R. M. Martin, *Intension and Decision: A Philosophical Study* (Prentice-Hall, 1963), chapter 2

[2] Ibid. pp. 45-46

6

There remains to indicate more explicitly how the notion of RPR relates to that of model. We assert—this is the second main point of the essay—that the metaphorical effect of viewing a theory through a model consists in carrying the RPR associated with the model over to the theory. This is to say that in addition to imputing the structure of the model to the theory in the sense of sections 2 and 3, the model user either consciously or unconsciously transfers the RPR associated with the *entire* model, including associations based on ' commonplace knowledge ' and on the unmapped sub-structure of the model, over to the theory. Thus the *formally* unimputed substructure of the model, along with a host of unformalised commonplace assumptions associated with it, acts as a filter—a metaphorical filter as outlined in section 5—such that the RPR that it determines is carried over to the theory to which the model is applied. In referring to a RPR ' determined by a model ', we mean no more than a RPR with respect to a class K of sentences employed by a typical user of the model. (A typical user of the planetary model would certainly prefer ' planets move in circular orbits ' to ' planets move in square orbits ', and ' planets move in elliptical orbits ' to both.)

A clear example illustrating the filter effect of an unimputed substructure can be got by adapting Goodman's example referred to in section three above. Interpret the ultimate factors K, L, M, N of a theory T_1 as points, and the complex elements replacing them in T_2 as pairs[1] of lines $a : 1$, $a : 2$, $b : 1$, and $b : 2$, according to the following diagram.



	1	2
a	K	L
b	M	N

Then the relation ' next-counter-clockwise-to ' among the points of T_1 will consist of the ordered pairs

$$\langle K , L\rangle$$
$$\langle L , N\rangle$$
$$\langle N , M\rangle$$
$$\langle M , K\rangle$$

and will be mirrored in T_2 by the ordered pairs

$$\langle a : 1 , a : 2\rangle$$
$$\langle a : 2 , b : 2\rangle$$
$$\langle b : 2 , b : 1\rangle$$
$$\langle b : 1 , a : 1\rangle$$

It is easily verified that on this particular replacement scheme, the pairs constituting each ordered pair in the relation ' next-counter-clockwise ' of T_2 share exactly one element in common (e.g. ' $a : 1$ ' and ' $a : 2$ ' have ' *a* ' in common). Our claim is just that this sort of ' excess ' sub-structure of models should, in general, be reflected in the RPR's associated with them, and should carry over to the theories for which they are models. Thus, to continue with the illustration, a preference (over its negation, say) for the sentence of T_2 ' next-counter-clockwise elements share a common member ' should be reflected, in an RPR associated with T_1, by a preference for the sentence ' K and L share something in common ' to *its* negation. Whether in actual practice such carry over leads to fresh insight culminating in new and confirmable (disconfirmable) predictions is, of course, an empirical question, unique to each particular case of model use.

---

[1] In the development of his example, Goodman uses unordered pairs ' $a : 1$ ', ' $a : 2$ ', etc., making it possible to talk about ' common members ' of two such pairs.

As for the expression 'carry over' used above, we shall only remark that although vague at this point of our analysis, it is *not* intrinsically metaphorical. It could be explicated and refined into a measure concept. That is the whole point of viewing metaphor as 'transfer', or 'carry over', of RPR's. The notion is essentially a behaviouristic and experimental one, susceptible to measurement, correction, and refinement. For suppose one finds a community of language users *x*, *y*, *z* (perhaps scientists) who display a certain common RPR_1 (call it their 'concensus RPR') with respect to a class K_m of sentences associated with (determined by) a certain model, the classical theory of planetary rotation, say. Suppose further that the model is used to explicate, or make more clear, a certain rudimentary theory of sub-atomic structure. The first thing that happens, of course, is that a certain obvious structure of the model is imputed to the theory. This is the standard isomorphism-of-structure part of model theory, although in view of our comments in section 3 we must recognise the asymmetry of the isomorphism. But now we expect something more. For in addition to the 'classical' imputation of structure from model to theory, we are now led to suspect that RPR_1—the concensus RPR of the model users—will be reflected in RPR_2, a concensus RPR, for the same users, with respect to a set of sentences K_t determined by the theory of sub-atomic structure. If 'planets don't normally collide' is preferred to 'there are frequent planet collisions' by concensus in RPR_1, then one would expect that 'electrons don't normally collide' will be preferred to 'there are frequent electron collisions' in RPR_2. Note that we are not claiming that this carry over is necessarily pragmatically useful. In some cases it is, in others not. It is merely analytic of 'good model' that the transfer *is* pragmatically useful, at least for some finite period of time. Recall that the Bohr-Rutherford theory, the one that actually made use of the planetary analogy, was vital to physics for only a fleeting moment, as historians see things. But if we cannot be sure that the transfer of RPR from model to theory is always pragmatically useful, at least we *can* measure the degree of transfer. Although we by no means wish to imply that the task would be simple, the general outlines of a program for such a measurement should not be too difficult to sketch. It might run in its stark outlines something as follows. Users of model and theory are interrogated by E with respect to the two sets of words constituted respectively by the theoretical vocabulary of the model and that of the theory, to determine which words pair off as analogues. (E.g. ‘electron’ with ‘planet’, ‘nucleus’ with ‘sun’, etc.) Given this particular Cartesian product of the two sets of theoretical terms, E could use it to construct, by uniform substitution, ‘analogue sentences’ in the theory from sentences in the model. Thus the model sentence ‘planets revolve around the sun’ would yield the analogue sentence ‘electrons revolve around the nucleus’ (we take ‘revolve around’ as a non-theoretical term of both model and theory). Suppose E has assembled k model sentences and their k analogues. There will then be ${}_kC_2 = \frac{k!}{2!\,(k-2)!}$ pairs of sentences,[1] for both model and theory, representing all the possible ways of taking any two of k sentences together. RPR_1 and RPR_2 could then be construed as the ranking given these various pairs of sentences. Now given a pair of sentences $\{a, b\}$ from the model, and the pair $\{a', b'\}$ consisting of their analogues from the theory, either they are scaled alike in RPR_1 and RPR_2 or they are not. A crude first measure of the heuristic ‘kick’ of a model might then be whether half, say, of the preference choices in RPR_1 match their analogues in RPR_2.

Not that we are actually suggesting that this crude measure of ‘transfer’ be taken seriously. What we *are* trying to underline, however, is that with a bit of ingenuity and perhaps some empirical feed-back, some measure or other of the carry over of RPR_1 to RPR_2 could be constructed. And once this were done, the notion of model would have sloughed off the aura of philosophical perplexity that has so long clung to it.

University of Massachusetts

---

[1] Recalling that the number of combinations ${}_nC_r$ of n things taken r at a time is $\frac{n!}{r!\,(n-r)!}$.

Brit. J. Phil. Sci., (1966), Volume 17, Number 4, 313–318
Printed in Great Britain

PARESIS AND THE ALLEGED ASYMMETRY BETWEEN EXPLANATION AND PREDICTION

PAUL DIETL

PROFESSOR Michael Scriven has recently suggested that the relation between syphilis and paresis offers material for a counter-example to the thesis that there exists a logical symmetry between explanation and prediction.[1] Professors Carl G. Hempel and Adolf Grünbaum have given independent arguments against the efficacy of the example. I wish to suggest in this note that the example does not show what it is intended to show but that the reasons Hempel and Grünbaum have given show a misunderstanding of the example rather than its impotence.

Here is Scriven's account of the example:

> We can explain but not predict whenever we have a proposition of the form ' The only cause of X is A ' (I)—for example, ' The only cause of paresis is syphilis '. Notice that this is perfectly compatible with the statement that A is often not followed by X—in fact, very few syphilitics develop paresis (II). Hence, when A is observed, we can predict that X is more likely to occur than without A, but still extremely unlikely. So, we must, on the evidence, still predict that it will *not* occur. But if it does we can appeal to (I) to provide and guarantee our explanation. . . . Hence an event which cannot be predicted from a set of well-confirmed propositions can, if it occurs, be explained by appeal to them.[2]

Grünbaum has argued that Scriven has failed to understand the nature of the asymmetry involved in the case.[3] According to Grünbaum, that which is involved does not invalidate the claim that there *is* symmetry between explanation and prediction. It is true, Grünbaum admits, that we are able to assert that syphilis *did cause* paresis while at the same time not being able to predict that syphilis *will cause*

---

* Received 25. v. 66

[1] M. Scriven, ' Explanation and Prediction in Evolutionary Theory ', *Science*, 130 (1959) [2] Ibid. p. 480

[3] A. Grünbaum, ' Temporally-Asymmetric Principles, Parity Between Explanation and Prediction, and Mechanism Versus Teleology ', *Philosophy of Science*, 1962, p. 165

paresis in any future instance but Scriven was wrong to think *this* asymmetry established a case against the symmetry Hempel argued for. Here is Grünbaum's argument:

> The sentences containing ' did cause ' and ' will cause ' [in his paper, paraphrased above, not in Sciven's] respectively *each* make *two* affirmations as follows: (i) the assertion of the *explanandum* (paresis) *per se*, and (ii) the affirmation of the obtaining of a causal *relation* (in the sense of being a contingently necessary condition) between the explanans (syphilis) and the explanandum (paresis). Thus, for our purposes the statement ' Syphilis *will cause* person Z to have paresis ' should be made in the form ' Person Z will have paresis *and* it will have been caused by syphilis ', and the statement ' Syphilis *did cause* person K to have paresis ' becomes ' Person K has (or had) paresis *and* it was caused by syphilis.' And the decisive point is that in so far as a *past* occurrence of paresis can be inductively inferred from prior syphilis, so also a future occurrence of paresis can be.[1]

Two questions arise: (A) Has Grünbaum correctly understood the example? and (B) Would the example, for all Grünbaum says against it, constitute a counter-example against the Hempelian thesis?

The answer to the first is, I think, negative. There is nothing in Scriven's account to suggest he doubts that we can have good evidence for the atemporality of necessity, no reason to think he would object to Grünbaum's proposition ' Precisely in the way and to the extent that syphilis *was* a necessary condition for paresis, it also will be! '[2] Therefore, the distinction between ' did cause ' and ' will cause ' is irrelevant. It follows that Grünbaum's translation of ' Syphilis will cause paresis in the future ' to ' Person Z will have paresis and it will have been caused by syphilis ' is equally beside the point. He has confused (*a*) ' because he has syphilis, he will have paresis ' with (*b*) ' if he develops paresis, he will have had syphilis '. This latter is not a prediction. Or if, you like, it is a conditional one.[3] Grünbaum's statement that future occurrences of paresis can be ' inductively inferred ' is mistaken.

---

[1] Op. cit. p. 166. It is important to notice that the phrases ' did cause ' and ' will cause ' are Grünbaum's and not Scriven's because putting things in those terms suggests the asymmetry is supposed to occur between our knowledge of future ones which, I shall argue, is just Grünbaum's confusion.

[2] Op. cit. p. 167

[3] Grünbaum's symbolism obscures the crucial distinction here. ' Z ' is not a proper name while ' K ' *is*. So in the future nothing about a person, Z, is being predicted just if anyone, Z, gets paresis he will have had syphilis. K, however, is a paretic whose condition we are actually explaining.

Speaking as if they could be predicted is to confuse necessary and sufficient conditions.

Nor is there any reason why Scriven should have denied the atemporality of the necessity here. Assuming A is a necessary condition of B then whenever B then A but not necessarily the other way around. Now if, in these circumstances, A counts as an explanans—*a view Grünbaum does not question*—then we do seem to have a cause of an explanans which even 'if taken account of in time, could [not] have served as a basis for predicting' the explanadum.[1] That we could quite correctly infer A from B now or in the future does not begin to show we could infer B from A. The answer to our second question is, therefore, that in spite of everything Grünbaum has written Scriven's example is unscathed.

If Grünbaum missed the difference between factually necessary and sufficient conditions, Hempel, in his answer to Scriven, missed the difference between factually necessary and logically necessary conditions.[3] Here is his argument.

> But precisely because the statistical probability for syphilis to lead to paresis is so small, and because on the given information we must rationally expect the given person not to have developed paresis, the information that the patient has had syphilis (and that only syphilitics can develop paresis) clearly no more explains the actual occurrence of paresis in this cause then a man's winning the first prize in the Irish Sweepstakes is explained by the information that he had bought a ticket (and that only a person who has bought a ticket can win the first prize).

But the cases are not parellel at all. Winning the prize is not just contingently connected with buying a ticket. Of course there may be other ways of getting prize money, e.g. you might steal a ticket from someone, or forge the winning number, but in all such cases 'winning the prize' is naturally put in quotation marks. It is just because buying a ticket in order to win a prize does not call to mind these devious means of getting the prize that we tend to agree with Hempel's claim that mentioning the purchase does not seem to have any explanatory power. It does not have any explanatory force because we treat it as a logically necessary condition. If it is not logically necessary, that is, if mention is made of the purchase in order to distinguish this case of winning from other ways one can come by the money, then it *is* explanatory, but then, of course, we are not dealing with a necessary condition of winning because we are not strictly speaking of winning but only of getting the prize (one way or another) and buying a ticket is not necessary for that.[1]

---

[1] Carl G. Hempel and Paul Oppenheim, 'The Logic of Explanation' reprinted in *Readings in the Philosophy of Science*, H. Feigl and M. Brodbeck, eds. (New York, 1953) p. 323

[2] C. G. Hempel, 'Reasons and Covering Laws in Historical Explanation' in *Philosophy of History*, S. Hook ed. (New York, 1963), pp. 148-149

[3] Ibid.

Actually I think the answer to Scriven is much simpler than the attempts by Grünbaum and Hempel would indicate. It can be drawn, in fact, largely from his own writings. The question is not whether or not necessary conditions are ever called causes. Sometimes they are and Scriven has undoubtedly found a case. The question, however, is whether explanations are ever possible when the explanans would not have provided a basis for prediction. We want to know not whether or not anybody ever called a necessary condition a cause but whether or not such a 'cause' would count as an explanation. Elsewhere, Scriven is quite explicit on the connection between explanation and understanding.

> Who is to say whether X and Y are understood? The *primary* case of explanation is the case of explaining X to someone; if there were no cases of this kind, there could be no such things as 'an explanation of X' in the abstract, whereas the reverse is not true. For it makes no sense to talk of *an explanation* which nobody understands now, or has understood or will, i.e. which is not an explanation for someone. In the primary case, the level of understanding is that of the person addressed.[2]

The question is whether mention of a necessary condition can ever 'fill a gap' in anyone's understanding. There are times when it would. If one of a group of apparently similar substances bursts into flame, you might explain it by mentioning that only that one had enough oxygen. That would be informative either to someone ignorant of what is necessary for combustion or to someone who does know but did not realise the oxygen was missing from the other samples. In any case, however, the necessary condition mentioned must be the only missing member of a jointly sufficient set; otherwise it would not explain why one pile burned while the others did not. But if it is, then prediction would have been possible.

---

[1] Some definitions of 'paresis' one finds in dictionaries *would* make syphilis a logically necessary condition, viz. those which define it as a late stage of syphilis or as a paralysis of syphilic origin (cf. the *O.E.D.*) but clearly this is not what Scriven had in mind. He was talking about 'paresis' in the general sense of cerebral paralysis and he assumed that syphilis, *as a matter of fact*, always precedes it.

[2] M. Scriven, 'Explanations, Predictions, and Laws', in *Minnesota Studies in the Philosophy of Science*, vol. 3. H. Feigl ed. (Minnesota, 1959), p. 205

There is one other circumstance in which mention of a necessary condition can be informative. It is the situation in which one (mistakenly) thinks the condition is absent and is puzzled as to how the effect could have taken place without it. It would ' fill a gap ' to inform him that there was oxygen there after all.

If neither of these situations obtain ' at the level of the person addressed ', then mention of a necessary condition cannot be informative. There are contexts in which any necessary or sufficient condition can be called a *cause*, but if we are both aware of the necessity of X, and of each other's awareness of it, then mention of X cannot be an *explanation*. If upon discovering my house burned down I turn to you in puzzlement and you tell me that there was the normal amount of oxygen in the neighbourhood, you have got to be making a joke.

Scriven's reply might be that if statements of necessary condition are not explanations then what is normally and properly called a causal statement may not be an explanation.

The answer, it seems to me, is to take seriously his own distinction between primary and extended uses of a term. If what I have said is correct, then giving a necessary condition cannot be an explanation in the primary use of that term. The primary use involves giving understanding to someone by mentioning either the sufficient condition (heating the gas increased its pressure) or the necessary condition other potential effects lacked (as in the combustion example). But no understanding can be given by mentioning a necessary condition of the explanandum unless someone is ignorant of or doubts the necessary connection. One might call these latter secondary uses of ' explain '. The point is that when syphilis is given as an explanation it cannot be given *as* an (undoubted) necessary condition. *Before* it was known to be a necessary condition of paresis, syphilis could, without oddity and in the primary sense, be given as the explanation (of some cases) of paresis. Once the necessary connection was established (or commonly accepted), however, although it is still called a cause, it can be seriously used to ' fill a gap ' only in the secondary sense of ' explanation.'

Now mention of *X* as an explanation of *Y* in the secondary sense amounts only to asserting X's presence (after all). No mention is made of the other conditions necessary but we do know they were there since *Y* did occur. If there is some evidence that some other condition, which we also had thought to be necessary, were not there then the how-possibly is not answered. 'How-possibly?', here,[1] clearly comes just to 'Which am I wrong about: (*a*) that *X* did not occur, (*b*) that *Y* did occur or (*c*) that *X* is a necessary condition for *Y*?' If it takes sufficient conditions to stop the 'how-possibly?' we need only assume determinism to give *X* (in those circumstances) predictive power. But *all* prediction rests on our belief that the future will resemble the past. Specifically, secondary explanations of events by mention of their non-logically necessary conditions are not different from explanation in general. They can be informative because they can, logically, be doubted. Hempel was wrong about that. They do not allow us to 'inductively infer' anything, however, because by themselves they are not even good evidence. Grünbaum was confused about that. But in the context in which we treat them as informative (not just refer to them as 'causes') we treat them as sufficient because in order to provide the understanding which a real explanation must, a causal explanation must provide sufficient conditions. That no prediction is possible is because we do not *know* all the others. That we acknowledge their presence is shown by the fact that mention of this condition stops the 'how-possibly?'. If we had taken account of all of other conditions we acknowledge, then the one we mention would have been predictive. Since what is required by way of assumption would allow a prediction of *Y* whenever *X* is informative about *Y*, I conclude that what Professor Scriven has written concerning the relation between syphilis and paresis does not constitute a refutation of the logical symmetry between explanation and prediction.

Temple University and
University of Illinois

---

[1] The case is more complicated in historical examples. Professor William Dray has introduced the example of how a fielder could possibly catch a ball flying twenty feet above the ground. The answer to this how-possibly is that he climbed a ladder. Now climbing a ladder is not sufficient and it is extremely difficult to say what *is* if for no other reason than the fielders do not always act up to capacity. No one has given any reason to think there is any comparable complexity which bears on syphilis and paresis however. See W. Dray, *Laws and Explanation in History* (Oxford, 1960), pp. 164-169.

NOTES AND COMMENTS

On the Principle of Total Evidence

Ayer (1957) raised the question of why, in the theory of logical probability (credibility), we should bother to make new observations. His question was not adequately answered in the interesting discussion that followed, by D. Bohm, R. B. Braithwaite, A. J. B. Cruikshank, P. K. Feyerabend, M. Fierz, W. B. Gallie, E. H. Hutten, W. C. Kneale, P. T. Landsberg, U. Öpik, M. Polanyi, L. Rosenfeld, M. Scriven, G. Süssmann, H. A. Thurston, and J.-P. Vigier. The question raised by Ayer is related by him to a principle called by Carnap (1947), ' the principle of total evidence ', which is the recommendation to use all the available evidence when estimating a probability. Ayer's problem is equally relevant to the theory of subjective probability, although, as he points out, it is hardly relevant to the theory of probability in the frequency sense.

In this note, Ayer's problem will be resolved in terms of the principle of rationality, the recommendation to maximise expected utility. (We use the words ' expected ' and ' expectation ' in the sense that is customary in nearly all books on mathematical probability or statistics.)

Our conclusion is that, in expectation, it pays to take into account further evidence, provided that the cost of collecting and using this evidence, although positive, can be ignored. In particular, we should use all the evidence *already* available, provided that the cost of doing so is negligible. With this proviso then, the principle of total evidence follows from the principle of rationality.

Suppose that we have r mutually exclusive and exhaustive hypotheses, $H_1, H_2, \ldots, H_r$, and a choice of s acts, or classes of acts, $A_1, A_2, \ldots, A_s$. It will be assumed that none of these classes of acts consists of a perpetual examination of the results of experiments, without ever deciding which of $A_1, A_2, \ldots, A_s$ to perform. Let the (expected) utility of act A_j if H_i is true be $U(A_j|H_i) = u_{ij}$. Suppose that, on some evidence, E, we have initial (prior) probabilities, $p_i = P(H_i|E)$. If just this evidence is taken into account, then the (expected) utility of act A_j is $\Sigma_i p_i u_{ij}$ and the principle of rationality recommends the choice $j = j_0$, the value of j that maximises this expression; and therefore the (expected) utility in the rational use of E is

$$\max_j(\Sigma_i p_i u_{ij}) = \Sigma_i p_i u_{ij_0} \quad .$$

We now consider making an observation whose possible outcomes are $E_1, E_2, \ldots, E_t$, where $P(E_k|H_i) = p_{ik}$ $(i = 1, 2, \ldots, r;\ k = 1, 2, \ldots, t)$. Let

$$q_{ik} = P(H_i|E.E_k) = p_i p_{ik} / \Sigma_i p_i p_{ik},$$

the final (posterior) probability of H_i if E_k occurs. (We denote logical conjunction by a full stop or period.) If in fact E_k occurs, then the expected utility of the use of E_k combined with E becomes

$$\max_j(\Sigma_i q_{ik} u_{ij}).$$

Now the initial probability of E_k is $\Sigma_i p_i p_{ik}$, so that the expected utility, in deciding both to make the new observation and to use it, is

$$\Sigma_k(\Sigma_i p_i p_{ik})\max_j(\Sigma_i q_{ik} u_{ij}) = \Sigma_k \max_j \Sigma_i p_i p_{ik} u_{ij}.$$

Accordingly we should like to prove that

$$\Sigma_k \max_j \Sigma_i p_i p_{ik} u_{ij} > \max_j \Sigma_i p_i u_{ij}, \tag{1}$$

with strict inequality unless the act recommended by the principle of rationality is the same irrespective of which of the events E_k occurs; in other words unless there is a value of j, mathematically independent of k, that maximises $U(A_j|E.E_k) = \Sigma_i q_{ik} u_{ij} = \Sigma_i p_i p_{ik} u_{ij}/\Sigma_i p_i p_{ik}$, or equivalently that maximises $\Sigma_i p_i p_{ik} u_{ij}$.

Since $\Sigma_k p_{ik} = 1$, the above proposition follows from the following Lemma by putting $f(j,k) = \Sigma_i p_i p_{ik} u_{ij}$.

LEMMA. *Let $f(j,k)$ be any real function of j and k. Then*

$$\Sigma_k \max_j f(j,k) > \max_j \Sigma_k f(j,k)$$

with strict inequality unless the matrix $\{f(j,k)\}$ has a 'dominating row'. (By a 'dominating row' of a matrix we mean a row in which each element is at least as large as any element in its own column.)

Proof of Lemma. Let a value of j that maximises $\Sigma_k f(j,k)$ be j_0. Clearly $\max_j(j,k) > f(j_0,k)$, since this would be true *however* j_0 were defined. The inequality is strict, when the definition of j_0 *is* used, unless $f(j,k)$ and $\Sigma_k f(j,k)$ are maximised by the same value of j. Therefore

$$\Sigma_k \max_j f(j,k) > \Sigma_k f(j_0,k) = \max_j \Sigma_k f(j,k).$$

This inequality is strict unless, for all k, $f(j,k)$ and $\Sigma_k f(j,k)$ are maximised at the same value of j. This establishes the Lemma and hence completes the resolution of Ayer's problem in terms of the principle of rationality.

At this point an opponent might say 'You have justified the decision to make new observations and to use them for the choice of the act A_j, but you have not justified the use of all observations that have already been made'. To this we can reply, 'The observations already made can be regarded as constituting a record. The process of consulting this record is itself a special kind of observation. We have justified the decision to make this observation and to use it, provided that the cost is negligible. In other words we have justified the use of all the observations that have been made, and this is the principle of total evidence.'

Our opponent might then say 'What you have shown is that, when faced with the two following possibilities, it is rational to select the second one:

(i) Not make an observation;

(ii) To make the observation and to use it for the choice of A_j;
but you have ignored a third possibility, namely

(iii) Make the observation and then not use it.'

My reply would be 'if we make an observation and then do not use it, this is equivalent to putting it back into the record. We have shown that it would then be irrational to decide to leave the observation in the record and not to use it, since there is a better course of action, namely to take it out (observe the record) and use it. You will now suggest other possibilities, such as the making of an observation, putting it on record, taking it out, putting it back, and so on, several times, and finally not using it. Our previous argument, with an obvious modification, shows that any such procedure is irrational, and it remains for you to suggest that your vacillating procedure should be continued for ever. But this would be a perpetual examination of the results of experiments, without a decision, and we have ruled this out by an explicit assumption.'

The simple mathematical theorem of the present note is not entirely new. Raiffa and Schlaifer (1961), p. 89, refer to the expected value of sample information, and seem implicitly to take for granted that it is positive. Lindley (1965), p. 66, explicitly states part of the theorem without proof. Perhaps the main value of the present note is that it makes explicit the connection between Carnap's principle of total evidence and the principle of rationality, a connection that was overlooked by seventeen distinguished philosophers of science.

Trinity College, Oxford I. J. Good

References

Ayer, A. J. (1957), 'The conception of probability as a logical relation', in *Observation and Interpretation*, ed. by S. Körner. London: Butterworths. Pages 12–30, including discussion

Carnap, R. (1947), 'On the application of inductive logic', *Philosophy and Phenomenological Research*, 8, 133–148

Good, I. J. (1950), *Probability and the Weighing of Evidence*. London: Griffin

Good, I. J. (1960), 'Weight of evidence, corroboration, explanatory power, information and the utility of experiments', *J. Roy. Statistical Soc.*, B, 22, 319–331

Lindley, D. V. (1965), *Introduction to Probability and Statistics*, Part 2. Cambridge: University Press

Raiffa, H. and Schlaifer (1961). *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University

I. J. GOOD

The White Shoe is a Red Herring

HEMPEL's paradox of confirmation can be worded thus 'A case of a hypothesis supports the hypothesis. Now the hypothesis that all crows are black is logically equivalent to the contrapositive that all non-black things are non-crows, and this is supported by the observation of a white shoe.'

The literature of the paradox is large and I have myself contributed to it twice.[1] The first contribution contained an error, but I think the second one gave a complete resolution. The main conclusion was that it is simply not true that a 'case of a hypothesis' necessarily supports the hypothesis; and an explanation was also given for why it seems to be true.

In the present note we show in six sentences, and even without reference to the contrapositive, that a case of a hypothesis does not necessarily support it.

Suppose that we know we are in one or other of two worlds, and the hypothesis, H, under consideration is that all the crows in our world are black. We know in advance that in one world there are a hundred black crows, no crows that are not black, and a million other birds; and that in the other world there are a thousand black crows, one white one, and a million other birds. A bird is selected equiprobably at random from all the birds in our world. It turns out to be a black crow. This is strong evidence (a Bayes-Jeffreys-Turing factor[2] of about 10) that we are in the second world, wherein not all crows are black. Thus the observation of a black crow, in the circumstances described, undermines the hypothesis that all the crows in our world are black. Thus the initial premise of the paradox of confirmation is false, and no reference to the contrapositive is required.

In order to understand why it is that a case of a hypothesis seems to support it, note that

$$W(H : \text{Black}|\text{Crow}) > 0,$$

where $W(H : E|G)$, the weight of evidence, support, or log-factor, for H provided by E given G, is the logarithm of the Bayes-Jeffreys-Turing factor, $P(E|G \text{ and } H)/P(E|G \text{ and not } H)$. The above inequality is clear from the fact that $P(\text{Black}|H \text{ and Crow}) = 1$, and a similar inequality will follow for all other explicata of corroboration.[3] On the other hand $W(H : \text{Crow}|\text{Black})$ can be negative.

It is formally interesting to recall that

$$W(H : \text{Black Crow}) = W(H : \text{Crow}) + W(H : \text{Black}|\text{Crow}),$$

and that only the last of these three terms needs to be positive. The first two terms can both be negative.

Trinity College, Oxford

I. J. GOOD

---

[1] This *Journal*, 1960, **11**, 145-149; and **12**, 63-64

[2] See either the first reference in note 1, or my *Probability and the Weighing of Evidence* (London, Griffin, 1950), chap. 6

[3] *Journal Roy. Statist. Soc.* **B**, **22** (1960), 319-331; and K. Popper, *The Logic of Scientific Discovery* (London, Hutchinson, 1959), Appendix ix

REVIEWS

Basic Concepts of Measurement. By Brian Ellis.
Cambridge University Press, 1966. Pp. 220. £2 5s.

Mr. Ellis has written a long-overdue book on concepts of measurement, which to some extent collects and unifies work he has published in separate articles during the last few years. He starts by considering the general notion of applied arithmetic, and then turns to those applications of it that are involved in measurement. He deals first with the concept of a quantity, then with that of a scale, and with the classification of scales. He goes on to propound a classification of measurement into fundamental, associative and derived, and draws from this an analysis of numerical laws, units, and dimensions. Finally he discusses the concepts of number and probability conceived as quantities. There follow a number of appendices, of which the most important is a translation of the passage in Mach's *Die Principien der Wärmelehre* dealing with the concept of temperature as a quantity. There is a good bibliography and an adequate index.

The most valuable feature of the book is its range and generality. There has been a good deal of disjointed work on the problems of measurement in particular sciences, into which distressingly unexamined general assumptions tend to be imported. Quite how a subject so fundamental has contrived to become treated in general as at once trivial and dull is one of the mysteries of current philosophy of science. Mr Ellis will deserve our especial gratitude for having rescued it from this state, as his book surely must. There is no doubt that it will start a considerable controversy, for his analyses certainly raise as many questions as they answer, as he would probably admit.

I think it must be a principal complaint against Mr Ellis that he does not push most of his analyses far enough. His explications are too vague and, having exposed a problem, he is too content to leave it with an appeal to such barely analysed intuitions as that of simplicity. Simplicity is a notoriously complex notion, which Mr Ellis treats far too imprecisely for his repeated invocations of it to be more than plausible. There is not space here to do more than justify by selective illustration the claim that Mr Ellis's valuable analyses are yet open to important criticism.

For example, in discussing the concept of quantity in Chapter 2, well rehearsed objections to Dingle's operationalism are produced somewhat repetitively. Mr Ellis's important point is that a quantity is defined by the order of items possessing it, not by any one of the different linear ordering relations that can generate the order. It follows that quantities, like other scientific concepts, are cluster (indeed law cluster) concepts. It is a pity that

Mr Ellis does not bring this more explicitly into the definition of ' quantity '. ' Quantity concepts are usually cluster concepts ' (p. 35) is too weak. We just do not admit quantities unconnected by our law network with other quantities.

Mr Ellis's other emphasis in this chapter, on the relational character of concepts, does not seem to me well placed. The general analogy with velocity is unsound. A single object A in the universe has no determinate velocity because there is no other sufficient set of objects, B, to which its motion can be referred, not because there are also no measuring rods and clocks, or no objects going faster or slower than A relative to B. To invoke such latter requirements in 'denying that it makes any sense to speak of a universe consisting of a single object possessing a variety of different quantities ' (p. 38) is to relapse again into operationalism, and to forget that a quantity is a disposition, which does not need to exhibit itself continuously or in any one particular way. Now it may be true that the possession of some quantity, e.g. the mass of a body, is determined by the distribution of other bodies in the universe, but if so, that is a quite different, contingent, fact, not at all supporting the claim that all quantities are necessarily relational.

The discussion in chapter 5 of the problem of why ' we choose the fundamental measuring operations that we do ' (p. 82) seems to me to be needlessly laboured. Mr Ellis admits that our notions of appropriate measures grow out of pre-theoretical practices, such as counting steps between places to measure distance. ' But this is only to explain our feeling [that the measure is appropriate], not to justify it ' (p. 82). What sort of justification other than explanation is Mr. Ellis looking for, given his own earlier criterion (of the correlation of different measuring operations) for the existence of a quantity measured? Surely the explanation affords a perfectly good justification for saying that what we now have is a measure of the pre-theoretical notion of distance. That is the natural consequence of agreeing with Mr Ellis's denial of a direct, absolute, measure-independent intuition of what a distance is, not that a further justification must still be given, through vague invocations of simplicity.

Mr Ellis's remarks in chapter 7 on the age of the universe have again a *prima facie* plausibility, but really need rather deeper analysis. In taking the questions,

> ' (*a*) Is there a theoretical maximum to the age of an event as determined on any given scale?
>
> (*b*) Was there an event before which no other event occurred? ',

to be separable, Mr Ellis must suppose that the term ' event ' can be used meaningfully independently of a choice of time scale on which events can be located. He may be right, but it is not at all obvious, and no arguments are presented to support such a view.

In chapter 10, one distinction between numbers and other quantities seems to me to be overdrawn. On page 155 Mr Ellis points out that the meaning of the sentence 'this group contains four things of such and such a type' is independent of counting operations. But equally, though perhaps not similarly, by Mr Ellis's own acceptance of quantities as cluster concepts, the meaning of 'this object is four metres long' is independent of any one particular method of length measurement. 'Our counting procedures are admitted as counting procedures, if and only if they give correctly the numbers of things in groups' (p. 155). Quite so, and the same is true of any proposed measuring procedure for length. No doubt with length it is a matter of correlation with other such procedures, which is perhaps not the case with counting. But anyway the point wants making more clearly than Mr Ellis makes it if he is not yet again to appear to be relapsing into operationalism.

Finally, despite his ingenious analysis, Mr Ellis does not persuade me in chapter 11 that there is a parallel between logical probability and a sort of 'logical' temperature. Let us follow recent literature in reserving the term 'chance' for the physical property which is the subject of statistical probability statements. Then the parallels to Mr Ellis's temperature statements,

T_i: 'The well-calibrated thermometer M has been immersed in the liquid L and the temperature reading t_i°C has been obtained' ($i = 1, 2, \ldots, n$) and

S_T: 'On evidence $T_1, T_2, \ldots T_n$, the temperature of L is $\bar{t}_n$°C' where $\bar{t}_n$ in some average of $t_1, t_2 \ldots t_n$.

are the chance statements,

C_i : 'A representative sequence of M trials on the coin-tossing set-up L has been made, and a proportion p_i of heads in it has been obtained' ($i = 1, 2, \ldots, n$)
and

S_C: 'On evidence $C_1, C_2, \ldots C_n$, the chance of heads in L is $\bar{p}_n$', where $\bar{p}_n$ is some average of $p_1, p_2 \ldots p_n$.

Mr Ellis asserts that S_T is an analytic statement of logical temperature on the strength of the analogy with S_C, which he must presumably take to be a statement of logical probability. But this is just false. In the first place, neither S_T nor S_C is analytic, and to suppose that they are is to forget yet again that quantities are cluster concepts and do not derive their meaning exclusively from one set of measuring operations. In the second place, S_C is not a statement of logical probability at all. The statement of logical probability that Mr Ellis is thinking of is

S_{PC}: 'On evidence $C_1, C_2, \ldots . C_n$, the probability of the hypothesis that the chance of heads in L is $\bar{p}_n$ is P_C.'

There is, of course, an exactly analogous statement of logical probability about the temperature statement S_T, namely,

S_{PT}: 'On evidence $T_1, T_2, \ldots T_n$, the probability of the hypothesis that the temperature of L is t_n°C is P_T'.

It is quite clear that there is no statement of 'logical temperature' even remotely analogous to S_{PC} and S_{PT}. What Mr Ellis has done is to show that *chance* has a logic of measurement and estimation exactly analogous to that of temperature—a conclusion with which I entirely agree. But he has not thereby provided the connection of chance with logical probability which he rightly criticises Carnap for neglecting. The required connection has been provided by Ian Hacking in *Logic of Statistical Inference* (C.U.P., 1965) with the so-called Frequency Principle (not to be confused with the straight rule of induction). It would be nice to have a justification for this principle (I suspect that Braithwaite may have inadvertently supplied it in 'Why is it reasonable to base a betting rate upon an estimate of chance?' in *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science*), but meanwhile it can be accepted as axiomatising a tacitly accepted connection.

These detailed criticisms of aspects of Mr Ellis's book are only made to throw its major, and pioneering, virtues into relief. It is a book well worth criticising. The central discussion of scales, of fundamental, associative and derived measurement is excellent, and one would like to hope that chapter 9 will put a stop to the flood of inane literature on units, dimensions and dimensional analysis. In all, a book to be read, critically and profitably, by all philosophers of science.

D. H. Mellor
Pembroke College,
Cambridge

Philosophic Foundations of Quantum Mechanics. By H. Reichenbach.
Cambridge (paperback reprint), 1965. Pp. x+182. 12s. 6d.

Professor Reichenbach's book on the quantum theory which was first published in 1944 is now available as a paperback reprint. It (i) gives an introduction into part of the mathematics of the elementary quantum theory; (ii) discusses the familiar problems of interpretation; and (iii) suggests a new interpretation in terms of a three-valued logic. I shall comment on the last two items only.

(ii) is based on the distinction between *phenomena* ('which are determined in the same sense as the unobservable objects of classical physics') and *interphenomena* (which can be introduced 'only . . . within the frame of the quantum mechanical laws'). It is assumed without further argument that the inferential chains leading to the interphenomena are 'of a much more complicated sort than the inferential chains connecting macro-occurrences with phenomena' (21). The reason is that 'in the inferences leading from macrocosmic data to phenomena we usually use the laws of classical physics' (21). This twin-assumption eliminates from the start all those interpretations which want to understand *both* phenomena and interphenomena in quantum mechanical terms (example: Ludwig's theory of measurement). It also shows that Reichenbach's famous 'principle of anomaly' is nothing but a trivial consequence of an arbitrary definition. For according to Reichenbach the principle asserts that 'the class of descriptions of interphenomena contains no normal system' (33). Now a normal system is a descriptional system for which (*a*) the laws are the same whether or not we make an observation; (*b*) the state of the object is the same whether or not we observe it (19). Also the definition of phenomena and interphenomena makes the distinction between the observed and unobserved coincide with the distinction between classical physics and the quantum theory (21). Hence, the principle of anomaly asserts only that the laws of the quantum theory are not the laws of classical physics. Some principle!

It is also pointed out that the attempt to give an 'exhaustive interpretation' of the phenomena leads to difficulties, or 'anomalies' as Reichenbach prefers to express himself. There are at least two non-equivalent explanations for what an exhaustive interpretation is supposed to be (cf. pp. 33 and 139) but we can safely assume that the one that is decisive for Reichenbach's argument ascribes well-defined values to all the classically definable magnitudes of an interphenomenon. A charged masspoint that has always a well defined position and a well defined momentum is an exhaustive interpretation of a specific interphenomenon (passage of a single electron). It is exhaustive interpretations of this kind which lead to the anomalies.

Such anomalies are: the sudden collapse of the wave accompanying a moving electron when it impinges upon a screen. The fact that particles in an interference experiment seem to 'know' what is going on at distant places (par. 7). Now obviously these 'anomalies' are but refuting instances of the assumption that the moving electron is a classical wave (first case) or a classical particle (second case). They show that for any theory which consists of the mathematical formalism of quantum mechanics together with some exhaustive interpretation there exist refuting instances. Reichenbach's attempt to eliminate them (which is what the whole book leads up to) is therefore, logically speaking, of the same kind as is the attempt to preserve a theory from refuting instances by a clever arrangement of ad hoc hypotheses. Let us now take a closer look at this attempt.

Reichenbach discusses four methods of removing the anomalies. *Method 1*: suggests that we should 'become accustomed' to them (37), i.e. it expects us not to be worried by the fact that a classical interpretation of the quantum theory turns it into a false theory. *Method 2*: advises us to use a certain interpretation only for describing those parts of the world where it works and to switch over to another interpretation as soon as a difficulty arises. The only difference between this interpretation and method 1 that the former uses two or more refuted interpretations whereas the latter uses one only. *Method 3*: suggests that we stop interpreting altogether and regard the statements of the theory as cognitively meaningless instruments of prediction (40). *Method 4*, finally, suggests that we change the laws of logic so that the statements showing the inadequacy of one of the chosen interpretations 'can never be asserted as true' (942). This is the method which Reichenbach himself adopts.

The objection against this method or, for that matter, against *all* the 'methods' which Reichenbach discusses is that they remove a clash between theory and fact not by suggesting new, and more adequate scientific theories, but by an interpretation of the existing theories that removes their faults from view. Of course, it would be too much to expect from a philosopher that he 'interfere with the methods of physical inquiry' (vii). But one should expect from him a little more than 'conceptually clarified' (vii) recommendations for the preservation of the *status quo*. Moreover, Reichenbach contradicts the criteria of adequacy which he himself introduces. According to these criteria, every law of quantum mechanics should have either the truth value 'true', or the truth value 'false', but never a truth value 'indeterminate' (160). Now such quantum mechanical laws or the Schrödinger equation, or the law of the conservation of energy produce, when coupled with an exhaustive interpretation, precisely the anomalies Reichenbach wants to remove. They must therefore all receive the truth value 'indeterminate'. (This follows from more formal considerations also.) Only meta-statements, such as the principle of complementarity which no one has as yet regarded as a physical law, are exempt from this argument. Reichenbach's suggestions therefore conform neither to the quite elementary demand to take refutations seriously, nor do they conform to the criteria he has set up himself.

Paul K. Feyerabend
University of California
Berkeley

Analytical Philosophy of History. By Arthur C. Danto.
Cambridge University Press, 1965. Pp. xi+318. 55s.

During the past few years there has been a marked tendency for philosophers interested in history to concentrate attention upon the notion of narrative. The story-telling aspect of historical writing has been, it is suggested, unduly neglected in favour of other matters less directly or intimately connected with the historian's central concern. In consequence, various important features of historical thinking and enquiry have been overlooked or distorted; for, as Professor Gallie pointed out in his recent *Philosophy and the Historical*

Understanding, ' it can be maintained that narrative is the form which expresses what is basic to and characteristic of historical understanding '.

Like Gallie, Professor Danto wishes to move the idea of narrative to the middle of the picture, exhibiting its rôle in history and indicating some of the ways in which this bears upon questions of the more traditional sort involving, e.g. explanation and objectivity. His approach to the topic is, however, a somewhat oblique one. Thus he begins by distinguishing between ' substantive ' (sometimes called ' speculative ') philosophy of history and the type of analytical enquiry in which he himself is engaged. The former seeks to give an account of the ' whole of history ', encompassing the future as well as the past—a misconceived and illegitimate enterprise. The latter, by contrast, tries to elucidate and explain, amongst other things, how the historian ' co-ordinates ' events in the past. This, though, presupposes that it is possible for historians to make meaningful and true statements about what has happened, and such a possibility has been philosophically questioned. The opening chapters of the book are therefore taken up with considering some of the forms this kind of secpticism may assume. Danto in fact singles out three. There is the claim that statements about the past are meaningless, because unverifiable: there is the claim that they may all be false, since there may have been (for all we know) no past for them to refer to: and there is the claim that every historical assertion is necessarily infected by a radical subjectivity deriving from the historian's personal attitudes and from his temporal and cultural position.

It is through his discussion of the last of these contentions that Danto is led to advance his own positive views concerning the significance of narrative. Historical relativism, as usually presented, tends to be vitiated by a number of unwarranted preconceptions. For example, history may be compared unfavourably with science, on the ground that (unlike the latter) it utilises ' overarching hypotheses or conceptions ' (Beard) which must somehow result in a fundamental misrepresentation or falsification of the facts; again, proponents of relativism often seem to presuppose some notion of a perfect or ideal account of the past which history as actually written invariably falls short of. Against such positions, Danto argues that the true distinction between history and science lies, not in the alleged circumstance that one uses ' organising schemes ' while the other does not, but in the very different *kinds* of scheme they respectively employ. It is the function·of history to ' tell stories ', and any narrative may be viewed as a ' structure imposed upon events '. As such, however, it will necessarily exclude certain events as being irrelevant and could therefore never conceivably satisfy the requirements embodied in the idea of a ' perfect account ' in which everything is taken care of and from which nothing is left out.

It might be objected that such an account could none the less be given by descriptions that did not use the narrative form; but here again Danto finds difficulties. Thus he claims that it is impossible to 'give a complete description of any event which does not use narratives', stressing the importance in this connection of a certain class of sentences ('narrative sentences') which 'refer to at least two time-separated events though they only *describe* (are only *about*) the earliest event to which they refer' (p. 143). In other words, it is a feature of descriptions of events using such sentences that they presuppose a knowledge of what occurred subsequent to the events they describe; they could not therefore be applied by contemporary witnesses in the absence of historical foreknowledge—a man writing in 1618, for instance, could not speak of the events of the year as 'the beginning of the Thirty Years War'. If—as Danto maintains—the requisite knowledge of the future is unattainable, the conception of an 'ideal account', together with that of a philosophy of history in the 'substantive' sense, must be abandoned. No limits can be set in advance concerning what future historians will make of the present or the past as these are known (or knowable) at any given date; and this is so, not merely because we cannot foresee the future course of events, but also because we cannot foretell the interests of future historians, and hence cannot anticipate the criteria of selection and relevance they will employ in the construction of their narratives. In so far as relativists have recognised the significance of such continuous 'retroactive re-alignment of the past' they have been right; their error has lain (it is implied) in lamenting, by reference to an impossible model, a situation which is in the nature of the case inevitable.

The above represents no more than the bare outlines of an argument which is developed at considerable length and with much complex elaboration; amongst other things, there is a thoughtful discussion of the logical structure of narrative explanation and of the kind of rôle such explanation plays in historical works. The exposition is not, however, always easy to follow, and it has puzzling features, not least in the sections dealing with determinism and historical foreknowledge. So far as the first is concerned. Danto seems partly to be contending that the existence of narrative sentences raises peculiar difficulties for determinists who believe any action to be predictable, given a perfect knowledge of its antecedent causes. Here I confess to not finding the reasons advanced very persuasive. Regarding the second, a number of considerations are adduced, but in such a way that it is left rather unclear whether it is the possibility of foreknowledge *as such* or whether it is the possibility of specifically *historical* foreknowledge that is being questioned. At one point Danto seems to be relying upon a quite general argument to the effect that, since 'propositions referring to the future are neither true not false', it follows directly that there can be no knowledge of them (pp. 196 ff.). Elsewhere, on the other hand, it is suggested that history presents special problems by comparison with other domains of enquiry; this is done partly by pointing out that any purportive prediction in the realm of human affairs is in principle open to falsification by relevant agents who are aware of the prediction, partly by casting doubt on the idea that there could be procedures for achieving the alleged knowledge in the historical sphere. The latter point raises the question of 'historical laws'. Danto admits that there might be such laws, but argues that, if discovered, they would not 'entitle us to conclude that there is historical inevitability any more than the existence of general laws of a non-historical sort entitles us to conclude that there is inevitability in nature' (p. 254). But, questions of inevitability apart, we certainly use general laws to predict natural events; and the problem arises as to why, given historical laws, the same should not be true of historical events as well. Danto, indeed, implies that—within certain limits—it might be true; his claim, however, that the discovery of such laws would 'in no degree' support the pretentions of substantive philosophy of history is less easy to understand.

Danto provides a salutary analysis of the ways in which the historian's conception and organisation of the past are dependent upon subsequent events and are hence inherently liable to constant alteration and modification; he stresses, too, how new preoccupations, fresh styles of thought and feeling, may involve radical revisions of previous interpretations, making it possible to read into what happened a significance to which earlier historians were blind. In this he has (as he notes) a valid point against some *Verstehen* theorists, who have spoken as if the historian's sole task lies in determining how things looked from the point of view of those who actually participated in the events he seeks to describe. Yet he seems at one place to lapse into exaggeration of the opposite kind; thus he writes (p. 183) that 'the whole point of history is *not* to know about actions as witnesses might, but as historians do, in connection with later events and as parts of temporal wholes'. The *whole* point? It is tempting here to quote Lovejoy's remark that it is imperative for the historian to realise that 'his ancestors had ends of their own' and that 'the content and meaning of their existence are not exhaustively resolvable into those of the existence of their posterity'. From what he says in general, however, I suspect that that is not an opinion with which the author of this interesting though at times perplexing book would want to disagree.

Patrick Gardiner
Magdalen College,
Oxford

Political Argument. By B. Barry.
Routledge and Kegan Paul, 1965. Pp. 364. 50s.

This book is a study of the relation between institutions and the principles by reference to which they are justified. It is intentionally limited in many respects. Firstly the principles chosen are those which have been prominent since 1945 in Britain and the U.S.A. Secondly it is concerned with 'justification' rather than 'evaluation'. Evaluation, Mr Barry argues, may take place without being able to formulate the precise principles on which a conclusion is based and it may involve ranking one principle above another without being able to give reasons. Justification, on the other hand, involves an explicit appeal to principles. It is obvious, therefore, why it is the proper object of study; for 'the limits of language are also the limits of philosophical analysis' (p. 3). There is also a further limit even in respect of Mr Barry's concern with justification. He distinguishes political principles (e.g. the maintenance of private ownership, commitment to neutrality) from 'ultimate considerations' such as justice and freedom. But he says little about the justification of such 'ultimate considerations' themselves. He assumes that they are plural rather than derivable from one principle and makes much of the interplay between such principles in his analysis of justification.

Mr Barry thinks (Chapter III) that the fundamental distinction between principles is whether they are 'want regarding' or 'ideal regarding'. He admits that such classifications 'reflect something about the author's outlook and what he regards as important' (p. 38). In his outlook the importance of want-regarding principles, or those 'which take as given the wants which people happen to have' looms very large, though why it should be thought that the satisfactions of the wants that people happen to have is of any ethical significance is not discussed. It is simply a 'right' (p. 42). Ideal regarding principles are characterised only negatively as those which refuse to accept each person's ranking of the importance to him of having different wants satisfied or, more radically, those which deny any value to the satisfaction of some wants, or, more radically still, those which attribute value to other things besides the satisfaction of wants—'to such things as people's tastes, characters, or beliefs' (p. 40). Within the total area of want-regarding principles Mr Barry draws a distinction between those which are distributive (e.g. equity and justice) and those which are aggregative (e.g. efficiency and expediency). This distinction is not altogether perspicuous. Freedom, for instance, which is interpreted reasonably enough as the principle that imputes value to the non-existence of human obstacles to want-satisfaction, appears as a major distributive principle. But what has this to do with distribution in any strict sense?

Chapter IV starts with a brief and rather thin critique of conservative theories such as those of Hayek, de Jouvenel, and Oakeshott, which hold that attempts to transform societies in accordance with principles are either pernicious or dangerous because of the tendency of their devotees to ignore the cost of pursuing them on other aspects of life no less important. After a discussion of majoritarianism Mr Barry deals with liberalism or 'the belief that one should try to implement politically only want-regarding judgments—never ideal-regarding judgments' (p. 94). This is one of the most interesting sections of the book. Mr Barry's distinctions do indeed illuminate the issues in this area of argument.

After a straightforward classification in Chapter V of types of social decision procedure (e.g. combat, bargaining, discussion on merits, voting, authoritarian determination) Mr Barry proceeds to the discussion of distributive want-regarding principles. His reason for concentrating on want-regarding rather than ideal-regarding principles is that he thinks that they are in almost universal use and therefore much easier to pin down than more vague and idiosyncratic appeals to ideal-regarding principles. He feels able to make general distinctions between 'justice' in the wide sense, which includes all distributive considerations, whether comparative or absolute, 'fairness' and 'equity' in a wide sense which indicate that some *comparative* distributive consideration is relevant, and two narrow applications of 'equity', one of which refers to modifications of general rules to meet special situations, the other to attempts to make precise comparisons within a general distributive pattern which is not itself questioned. Whatever is to be said about the distinctions made by Mr Barry between these highly indeterminate concepts, his discussion of forms of argument in this area is extremely stimulating. He is at times rather cavalier in his comments on others. For instance (to cite an example which requires little research to estimate the justice of his comments!) he accuses Benn and Peters of 'emasculating' the concept of equality by formulating it as the demand that reasons must be adduced to justify treating people differently, without taking account of the fact that their formulation is to be seen in the context of an attempt to justify the general notion of 'justice'. Mr Barry has no interest in the problem of justifying what he calls 'ultimate considerations'; but he should not call a formulation subservient to such a preoccupation a 'hostile' one. His discussion is of particular interest when he admits the possible intrusion of ideal-regarding principles—for instance in the case of issues to do with integration and non-discrimination (Chapter VII), in which he raises the question whether integration is a value on its own or not. His discussion of ideal-regarding forms of argument for freedom (Chapter VIII) are not so interesting because he assumes that these must have something to do with people being made 'better'. He never asks whether the justification of freedom might depend on its status as a presupposition of practical discourse.

The rest of the book is concerned with aggregative principles. It rests, to a large extent, on a naturalistic definition of 'interest'. An action or a policy is 'in a man's interest' if it increases his opportunities to get what he wants. Mr Barry considers several minor objections to this account but gives no serious consideration to the 'ideal-regarding' approach which would reject outright this analysis of 'interests' in terms of what people actually want. A policy is often rejected as being in a person's interest, not because it would not in fact tend to maximise a person's opportunities for getting what he wants, but because he wants all the wrong things. Mr Barry's attempt to answer Mr Benn's point, that 'acting in the interests of a child' does not depend necessarily on having regard to what the child wants, rests on superficialities. It is true that this phrase occurs in a special context; but *all* considerations of what is in a person's interest occur in special contexts. The argument that schools and parents are concerned with 'education' or 'rearing', which are different notions from 'acting in a person's interest', is valid in so far as there is an obvious distinction between such notions, but beside the point; for schools and parents are *also* concerned with what is in children's interest, and there may well be a connection between 'education' and 'considering people's interests' in that *both* are related, though in different ways, to some conception of a worth-while form of life.

Mr Barry's analysis of the basic concept of 'interest' is inadequate in many other ways. He never explores, for instance, its relation to notions like that of 'interesting' or 'interests' in the sense of 'hobbies'. But his analysis of derivative notions such as 'the public interest' and 'common interest' in Chapter XI and of applications of 'the public interest' in Chapter XII is much more deft. Indeed it is one of the most illuminating parts of the book. Chapter XIII contains a very good examination of traditional arguments for claiming the peculiar appropriateness of public interests as matters for the concern of the state.

The book concludes with two chapters which discuss the central perennial concern of political theory with the problem of how to ensure that the force of the state will be applied to the pursuit of the public interest and not to some quite other purpose. After a discussion of the power concentration theory a highly sophisticated version of the power-diffusion theory, put forward by Buchanan and Tullock in *The Calculus of Consent*, is criticised in considerable detail. The question is raised in the concluding sections whether an assumption of universal self-interest is an appropriate psychological theory upon which to raise a political theory. The book ends on a negative and rather unsatisfactory note with a warning about the danger of supposing that such highly complex questions could be settled by the appeal to general statements or formulae.

Mr Barry's overall intention in writing his book has been to demonstrate the importance of analytical philosophy in the discussion of political principles, to encourage the fusion of analytical philosophy with analytical politics.

> The analysis of concepts becomes beyond a certain point tedious logic-chopping; even more, the attempt to draw practical conclusions from political models can only too easily come to grief on excessively crude (or—even worse—vague and evasive) value premises (p. 291).

Many may regret his failure to raise even more fundamental questions about the justification of his 'ultimate considerations'; many may criticise his analysis of particular concepts; but any fair-minded reader must applaud both Mr Barry's intention and the admirable competence and erudition with which he has executed it.

R. S. Peters
University of London

REVIEWS

The Stratification of Behaviour. By D. S. Shwayder. Routledge and Kegan Paul, London, 1965. Pp. xvi+411. 56s.

DR. SCHWAYDER has written a large and complex book. It is, moreover, according to his avowed intentions, but part of what was to have been a comprehensive work on the Theory of Language, but which became too large to constitute a single book. Shwayder speaks in the preface of a 'separate volume' and later of 'subsequent volumes'—a thought which may well be found intimidating. The present book is large and complex enough; its origins are apparent in the way in which it leads up to discussion of language as a form of action, breaking off finally and somewhat incoherently in some hints concerning what is to come. Shwayder shows awareness of its excessive length, and claims even to have compressed the presentation; but it nevertheless a book ruined by length—perhaps indeed a paradigm case of such. Shwayder prefaces each section with a summary of what is to come—a procedure which may be of help to anyone trying to find and remember his way through the maze of points on the philosophy of mind and action which Shwayder throws out.

Apart from the length, there are further difficulties for the reader in the method. Shwayder claims to have been more influenced by Austin than by anyone else, and this is evident enough in particular details. But it is doubtful whether the spirit of the book is Austinian. Its sub-title is 'A system of definitions propounded and defended', and this is a fair indication of the approach. Shwayder is chiefly concerned to set out and elucidate a whole range of concepts necessary to and useful for an analysis of human action. This involves him in classifying and distinguishing forms of action and behaviour, etc., in a technical way. Some of these technical distinctions correspond in some ways to points which have been made by other philosophers, but Shwayder is concerned to present them systematically; hence the technicality. What is less clear is the extent to which such a programme can really be carried out. For, Shwayder is not presenting a technical scientific theory of behaviour, but a philosophical theory, the aim of which, presumably, is to increase our understanding of just what behaviour is and can be. Any systematic theory of this kind has to start from somewhere—presumably in our ordinary understanding of the notions involved. But this does not appear to be Shwayder's procedure, for he offers at the outset a technical definition of 'item of behaviour' and then defines 'action' in terms of this together with a technical concept of 'purpose'. Into the initial definition of 'item of behaviour' is written a philosophical thesis about responsibility, i.e. that behaviour consists of animal movements with the presumption that the animal in question could, in certain circumstances, have 'arrested or inhibited movements of the kind in question' (pp. 21 and 26 ff.). This, he says, is a refinement of the avoidability criterion of responsibility. But surely this is normally offered as a criterion for our *ordinary* notion of responsibility, and it is here in any case that we should require argument. How does it fit in with an avowedly technical notion of ' behaviour '? It seems as if Shwayder wishes to present technical concepts in a systematic way but to slip back into non-technical considerations when it suits him.

The same sort of thing is evident in his treatment of ' purpose '—the thing which distinguishes action within the wider class of behaviour. It is once again an avowedly technical concept. For reasons which are not altogether clear, Shwayder objects to the more fashionable notion of intention, as used by Anscombe and others. But his notion of ' purpose ' is not very far from this. It is, he says, what is specified in answering the question ' What was the animal meaning to do?' To give the purpose of an action need not involve going beyond the action itself, and every action has a purpose, there being certain criteria for the presence of purpose in behaviour—(1) the possibility of success or failure, and (2) the possibility of the behaviour's being explained by the animal's knowledge or beliefs. Once again, one is at a loss to discover whether this is *all* technical or whether it has some relevance to our ordinary concept of action. And so it goes on. There is a constant tendency to meet possible objections by saying that the notion under consideration is technical, and as a consequence of this it becomes less and less clear what is achieved by the discussion. Moreover, many of the discussions of ' received ' philosophical issues are tangential, e.g. on knowledge without observation (p. 80) and the theory of descriptions (pp. 383-384), to mention but two instances. A further result of all this is an inevitable appearance of dogmatism on many issues. Shwayder claims to be using a kind of species-genus method of classification of the subject matter, but the criteria of validity for the method in this context are never adequately explained or justified.

I have space to comment on only one point of substance—but it is, I think, an important one. In his treatment of rules, Shwayder rightly distinguishes rules from generalisations and insists on the priority of what he calls community rules to private rules. But he also maintains that rules are systems of expectations—the idea being ' that one conforms when he acts with the reason or mistaken reason that others are entitled to expect him to act so ' (p. 257). Even if this is necessary to conformity to rule, I do not think that it is sufficient. Indeed, I think that there is here and elsewhere in the book a serious under-estimation of the role of social norms, which is also to be found in Grice's paper on ' Meaning ' to which Shwayder refers as a major influence. This is not without significance.

On many issues, some perhaps incidental, Shwayder has wise remarks to make, and the comprehensiveness of the approach is impressive, even if it is, I suspect, misguided. But the scale of the work and the method followed tend to leave one with no clear impression of what has been achieved. It is on incidental issues, perhaps, that there is most here to be learned.

D. W. HAMLYN
Birkbeck College, London

REVIEWS

The Pursuit of Certainty. By Shirley Robin Letwin.
Cambridge University Press, 1965. Pp. vii+391. 52s. 6d.

SHIRLEY LETWIN has written a large, expansive, wide-ranging, elegantly produced and prettily written, non-book. Inside it there are several excellent articles and a book or two struggling to get out. The whole is titled, promisingly enough, *The Pursuit of Certainty*; the parts, more precisely and more illuminatingly, '*David Hume*', '*Jeremy Bentham*', '*John Stuart Mill*', '*Beatrice Webb*'. So far, so good. One might regard the whole as promising something more than the sum of the parts—the expectation is not wholly disappointed—but the sum of the parts would in itself represent a substantial total. None of the four figures has been exhaustively treated by historians and—except perhaps in Halévy's *La Formation du Radicalisme Philosophique*—none of them has been discussed at length as political writers in a profoundly illuminating way.

On all of them Mrs Letwin casts considerable light. Hume as the British Montaigne; Bentham as the political gadgeteer; John Stuart Mill as the strenuous devotee of 'the Religion of Humanity'; Beatrice Webb as the narrow relentless intellectual who mistook her own self-assurance for science and took the Russia of Stalin to her improbable heart; all are set off effectively, are made to take their bows and disclose their moral. The portraits are painted with considerable psychological skill and an impressively broad learning. The picture of Bentham in particular is sharply focused and that of Hume almost verges on idolatry. In both cases, it is not so much that something altogether new has been glimpsed as that something which we knew before is seen forcefully and presented with great clarity. She is perhaps, both less charitable towards, and less perceptive about, John Stuart Mill and Beatrice Webb. Indeed the latter is transformed into a formidable ogress. But even here something real and instructive is brought out.

Several questions, however, remain. Why this sort of book? For what group of readers was it written? Most crudely and disturbingly, what is it supposed to be about? They are, of course, all different ways of asking the same question. But the difference in focus makes it easier to suggest the character of the puzzlement one might feel. To take the section on Hume as an example—it is quite long—more than a hundred pages—and it discusses many features of Hume's attitude to politics with real sympathy and discernment. But the whole effect is curiously off-hand. There are, after all, three extended treatments of Hume's political theory which have been published within the last twelve years. One of them, John B. Stewart's *The Moral and Political Philosophy of David Hume* (New York, 1963), might conceivably have appeared after Mrs Letwin's book went to press. But the other two, Giuseppe Giarrizzo's *David Hume Politico e Storico* (Turin, 1962) and Georges Vlachos's *Essai sur la Politique de Hume* (Paris, 1955), can hardly have failed to come her way because of this. Yet at no point is so much as a nod made towards them. There might in some cases be good reason for this sort of attitude. The whole previous tradition of interpretation might be so misleading that it is simpler and clearer to ignore it, in exposition. Or the books in question might be so bad that nothing of value could be learnt from them. Even so, some perfunctory acknowledgment of their existence might seem natural. And in practice neither of these possibilities seems to apply. Mrs Letwin writes incomparably more elegantly than Stewart or Vlachos (and somewhat less oddly than Giarrizzo) but she does not make much clear that was not clear already. If what is intended is an elegant portrait of the Humean attitude to politics, it seems curiously lengthy and punctilious in its presentation. Yet if what is intended is an adequate presentation of Hume's thinking about politics, as he conceived it and modified it over the years, then it seems in many ways inadequate and any attempt to present it at this length and in this manner must necessarily have been so. As a set of discrete studies, the book seems in this way to fall uneasily, though not ungracefully, between two stools.

But, as the title makes clear, it is not intended as a set of discrete studies. It has a 'theme'. The 'theme', though, is not very easy to identify. At times, it seems to be the story of how the British political tradition got corrupted by the erroneous opinion that one could make scientific statements about politics. At others, it seems more psychological; a story of how three 'representative' figures became corrupted by their need to hold unchallengeable political opinions (either way what a *horrid* descent from David Hume to Beatrice Webb.) The first theme is a grand and complicated one; but its grandeur rather depends upon its being true that one cannot make scientific statements about politics. Such a proposition clearly demands some argument (which it does not receive). If, on the other hand, it is *not* a premise of Mrs Letwin's book, her approach becomes harder to understand. If the story is *not* a funny story (about the attempt to create a science of political astrology), if it is more aptly seen as studies in the pre-history of a science, why is so little detailed attention paid to the concrete enterprises of the 'scientists'? The most sustained piece of argument in the book, the chapter on the sociology of John Stuart Mill is also in some ways the most rewarding (it contains one marvellous joke), not because it is more convincing than most of the other chapters, but because what Mrs Letwin is saying in it is much clearer than elsewhere and this clarity makes for more interesting perspectives on what Mill was saying.

But perhaps this is all beside the point. The Introduction to the book has a very distinctive tone and it may be here that its theme really lies. It is subtle, complex, ironical, a little reminiscent of Michael Oakeshott.... '... we are forcibly reminded that names and doctrines collect a variety of associations and hide a limitless stock of temperaments, beliefs, and purposes. These are arranged in unforeseeable patterns. And neither affinities nor divisions, in politics as elsewhere, are ever complete or simple'. 'Forcibly' perhaps does not do justice to the delicacy of Mrs Letwin's style; but we certainly are reminded. The pursuit of certainty may have corrupted Beatrice Webb into reducing tragedy to inefficiency, but it does not seem to have overwhelmed the author. This book is a lot more sophisticated and a lot more poised than most academic books about the history of political theory. It is the impressive product of a considerable intelligence and a wide learning. But it does not add much, crudely, to knowledge—not even to knowledge as coherent and well-established interpretation. The scope of the ambition makes such definition almost impossible. And yet, for all the poise of the writing and the pleasure it will give to readers, some sharper definition is much needed. There are costs to pursuing elegance of manner before certainty. Perhaps in the end it is an open question which way the irony of the title turns.

John Dunn
Jesus College, Cambridge

Philosophische Bemerkungen. By Ludwig Wittgenstein. Ed. R. Rhees. Blackwell, Oxford, 1964. Pp. 348. 35s.

1. *Historical remarks*. When Wittgenstein returned to Cambridge in 1929 he did so largely under the pressure of his English friends. This fact, however, does not yet explain his return to philosophy. As one can see clearly from his extensive *Philosophische Bemerkungen* (*PB*) written in 1930 this return to philosophy was certainly not motivated by a radical break with the ideas of the *Tractatus* (*TLP*); it may have been the result of a feeling of uneasiness with the doctrines of the early work. This uneasiness may in turn have been due to the influence of Schlick and Waismann on the one hand and that of Brouwer on the other. It has been said that, in particular, a lecture by Brouwer delivered in Vienna in 1928 (and published under the title 'Mathematik, Wissenschaft und Sprache' in *Monatshefte für Mathematik und Physic*, 1929, 36, 153-164) stimulated Wittgenstein to take up philosophy again. The extensive discussion of mathematical problems in *PB* suggests that this is, in fact, a reasonable assumption to make.

2. *Topics from TLP*. About half of the book is taken up with discussions of topics which are well known to the reader of *TLP*. It is, of course, not possible to discuss all of them; therefore, three items of particular interest will be selected.

(*a*) In *PB* Wittgenstein contrasts his picture theory of meaning with Russell's theory of meaning. Wittgenstein characterises the difference by maintaining that for Russell the recognition that a given fact makes a thought true is something over and above the thought and the fact, while according to the views of *TLP* the relation between the thought and its corresponding fact is an internal relation. This supposed internal relation between thought and fact is, presumably, the 'sense' of the thought. In section III of *PB* Wittgenstein uses the term 'intention' and says: 'If the element of intentionality is removed from language, language thereby loses all its function.'

(*b*) At this point it becomes clear what Wittgenstein means by the assertion: 'Every description can be regarded as a prescription' (para. 14); for every description can be regarded as the instruction to find the fact corresponding to the description or to form a picture of such a fact; for 'the essence of the intention is the picture' (para. 21). It follows that 'strangely enough the problem of the understanding of language has something to do with the problem of the will' (para. 13). Though Wittgenstein also suggests that every prescription can equally be regarded as a description there can be little doubt that his thought is at this point preparing for a development which, in the end, led to the views contained in the *Investigations.* While *TLP* considers the description of a simple state of affairs as the paradigm case of a meaningful sentence, *PB* at least suggests a wider point of view. These remarks from *PB* are, in fact, reminiscent of the intuitionist philosophy of language as expounded by Brouwer in his 1928 lecture. According to this theory the paradigm case of a meaningful sentence is in fact the command and not the description. Perhaps we can take the considerations of *PB* as an attempt to reconcile the views of *TLP* with those of Brouwer.

(*c*) In the last section of *PB* Wittgenstein discusses the question in what way we come to adopt or reject propositions of science. His attitude here seems to be considerably more pragmatic than in *TLP*. At one point he goes as far as to claim: 'The hypothesis stands in a looser relation to reality than that of verification' (para. 227). Wittgenstein is of course aware of the fact that if he were to combine this more positivist attitude towards scientific hypotheses with the assumption that such hypotheses are meaningful sentences in the sense of *TLP*, this would bring about the immediate collapse of the picture theory of meaning. He therefore, presumably, claims that a hypothesis is not itself a sentence, but merely 'a law for the formation of sentences' (para. 228). But even this move must tend to the destruction of the picture theory, since it now follows that a large number of what would ordinarily be called sentences are not in the strict sense meaningful sentences at all. It seems reasonable to assume that these attempted modifications of the picture theory of meaning were due to objections which the members of the Vienna Circle had raised against that theory, and which presumably were brought to Wittgenstein's notice through his discussions with Schlick and Waismann.

3. *Mathematical problems.* Though the influence of Brouwer's thought is clearly to be felt in the discussion of mathematical problems contained in *PB*, Wittgenstein's own views by no means coincide with those of Brouwer. They rather tend in the direction of finitism and formalism. In fact, it

appears not exaggerated to say that Wittgenstein's views on the nature of mathematics underwent much less change than his views about language; for what he has to say on this topic in the *Remarks on the Foundations of Mathematics* appears not to be essentially different from what he says in *PB*. In view of this it is perhaps less surprising that, in spite of the obvious stimulation by Brouwer, *PB* presents no serious examination of mathematical intuitionism.

Most unsatisfactory are, however, Wittgenstein's comments on set theory. In *TLP* 6.031 Wittgenstein says: 'The theory of classes is altogether superfluous in mathematics.' This statement has often been regarded as a mere endorsement of Russell's no-class theory. Remarks we find in section xv of *PB* however suggest a much stronger interpretation. In paragraph 173 we read: 'Mathematics is completely contaminated with the pernicious set-theoretical terminology.' And in paragraph 174 Wittgenstein says: 'Set theory is false because it pretends to presuppose a symbolism which, in fact does not exist, instead of the one which exists (which alone is possible). It is erected on a fictitious symbolism, and hence on nonsense.' It is in this hostile attitude to set theory that, perhaps, one can find the source of much that is doubtful in Wittgenstein's philosophy of mathematics. Though Wittgenstein's views are obviously connected with those of Hilbert and Brouwer, he completely failed to see why neither of these two critics of classical mathematics completely rejected set theory. He never even seems to have seen the reasons why Hilbert developed his programme for the justification of non-finitist methods, or why Brouwer introduced into mathematics the notion of rule-governed and freely proceeding infinite sequences. Both these steps are of course designed to deal with what has been described as 'a central, presumably even *the* central, problem of the foundations of mathematics' (Fraenkel), namely, the problem of an analysis of the continuum. It cannot be said that Wittgenstein was seriously aware of this problem in *PB*.

PB is in many ways a highly inconclusive and thoroughly irritating work. What Wittgenstein has to say in it about the philosophy of mathematics is of very limited value to actual discussions in that field. Perhaps the most satisfactory way to read *PB* is to read it as a critical commentary on *TLP*.

H. D. Sluga
University College London

The Voices of Time: a cooperative survey of man's views of time as expressed by the sciences and by the humanities. Edited by J. T. Fraser.
Braziller, New York, 1966. Pp. xxv+710. $12.50.

It is said that Arabic has more than a hundred words which may be translated 'sword'. Having read this volume, I am astonished that we have been content for so long with one word for the many concepts of time. It must be presumed that philosophers of science, who are notorious coiners of words, all believe that they have laid Time by the heels. I am not aware of any previous occasion on which such a large number of authors (twenty-six in all) have been commissioned to do precisely this, simultaneously. Many of the contributions are of the highest quality. There are also fifteen short literary pieces, by the editor, scattered more or less at random, which are not completely successful at turning this into a bedside book. It is well arranged, even so, and one of its chief merits is its encyclopaedic quality, enhanced by a wealth of bibliographical reference and an excellent index. (Librarians need read no further: they should buy the book at once.)

There are four sections. The first, *Time in Thought*, serves to remind us of the attitudes to time held by thinkers outside the narrow world of the academies. These may, for example, be investigated historically, as by J. L. Russell, S.J. ('Time in Christian Thought') and S. G. F. Brandon, or linguistically, as by Hajime Nakamura ('Time in Indian and Japanese Thought'). Joseph Needham unites the two approaches in a masterly way, despite some controversial asides on the early history of mechanical time-keeping.

The philosopher, intent on abstracting the time concept from acceptable scientific theories, finds it all too easy to overlook the importance of the individuals' whole experience of time. Recent psychological investigations are well summarised in John Cohen's contribution, and between this and the other five essays in the second section (*Time and Man*), explanations are given for the effects which hope, anxiety, forebodings of mortality, and so on, have on our time-awareness. Part three—*Time and Life*—follows with five essays in which relevant biological work is considered. Hudson Hoagland provides a bridge between the two parts, in proposing a biochemical explanation of the human time sense (in terms of the rate of oxidative metabolism in the brain). As a bonus, he has something to say on the possibility of keeping human sperm in suspended animation for an indefinite period. The concept of fatherhood, it seems, is due for revision.

The largest section of this compilation is the last—*Time and Matter*. The history and technique of time-keeping is dealt with by H. A. Lloyd and G. M. Clemence, who between them give a fairly accurate picture. (The editor makes clear the reason for the inclusion of these essays: 'Be that as it may, the affinity of clocks to time is generally admitted'.) Milič Čapek gives a spirited account of the status of time in special relativity theory. Roughly speaking, a contribution by O. Costa de Beauregard (the 'Philosopher of Being') supports the now famous dictum of Minkowski: 'Henceforth space itself, and time itself, are doomed to fade away into mere shadows, and only a sort of union of the two will keep an independent reality.' The reason: the individual can never claim to have divided up events absolutely and unambiguously into past, present and future, since observers moving relatively to them will make a different division. Time and space are thus merged, and on an extreme view it is even said that *change* (which requires a lapse of time) is impossible.

Čapek (the 'Philosopher of Becoming'), points out that the *individual* can make this unambiguous division. On *his own* world line, the Now moment separates past from future absolutely on that world line, and this separation holds (for events on that world line) in all other possible frames of reference. The emphasis here is on the individual and his experience. The Here-Now-for-Me takes over from the Newtonian Everywhere-Now.

Herbert Dingle ('Time in Relativity Theory: Measurement or Coordinate?') emphasises the irrelevance of time-as-experienced for the theory of relativity: the theory 'is wholly concerned with "time at a distance"'. This does not, of course, mean that he agrees with either de Beauregard or Čapek, for, as is widely known, he believes Einstein's special theory to contain an inconsistency. This is explained with all his usual clarity (pp. 466-467). Briefly, for those who have not encountered it before, it is shown on Einstein's principles that one of two clocks, both running regularly, may show a longer lapse of time than the other, between one pair of events, and a shorter lapse between a second pair. The one clock would be said to be running both faster and slower than the other. 'This is a plain contradiction.' It is nearly a decade since Dingle first tried to get a fair hearing for Ritz's ballistic theory of light emission. The resulting controversy has tended to osbcure the supposed contradiction in Einstein's theory. Is it a contradiction? Perhaps the time will come when writers of textbooks on relativity will decide otherwise, and place 'Dingle's denial of absolute clock-rate' alongside 'Einstein's denial of absolute simultaneity'—not that the controversy will then cease.

E. J. Zimmerman considers the message of quantum theory for a philosophy of time, without feeling compelled to draw any firm conclusions. Richard Schlegel has little new to say on time and entropy, but his treatment of the subject is clear and concise, and his essay could be added to any undergraduate reading-list. Satosi Watanabe's 'Time and the Probabilistic View of the World' is much sterner stuff, the second longest and perhaps the most closely argued contribution to the volume. His thesis that the direction of entropy increase in our experiments

> 'is determined by our special way of preparing the initial state and this is not in nature itself but originates from the side of the experimenter'

deserves a review to itself.

In the last essay, G. J. Whitrow elaborates upon some of the cosmological parts of his *Natural Philosophy of Time* (1961). I hope I am complimenting both in saying that, in most respects, *The Voices of Time* provides a worthy supplement to Whitrow's book.

J. D. North
Oxford University

Science and Philosophy. By Vincent Edward Smith.
Milwaukee: Bruce Publishing Company, 1965. Pp. xiii+266.

POSITIVISM is wearing thin, purpose has been reintroduced in biology, and physicists are rediscovering primary matter. These are the reasons alleged by the author to revitalise and refurbish what this reader takes to be a brand of Thomism.

Well, we can always take a skeleton out of the philosophical cupboard. The question is, will it still rattle? Mr Smith is quite convinced that it will not, and the upshot of his argument is to be found in chapters i, vi and vii. Chapters ii and iii provide a thumb-nail sketch of the history of science. The remaining chapters iv and v review, evaluate, and conveniently label philosophers from Plato and Zeus of Citon to Whitehead and A. J. Ayer.

Mr Smith warns us of the danger of oversimplification and the warning should be heeded. His outline of the development of science from Ptolemy to the twentieth century tends to create the impression that science is something that grows inevitably from one stage to the next without alternatives, by-roads, or superfluous appendages. We are told, for instance, that the 'law of inertia, though formulated earlier in the history of thought, was given strong confirmation by Galileo' (p. 48). Now the historically interesting and philosophically significant fact is that Galileo *did not* discover the law of inertia in spite of the mental experiments that *should* have led him to find it. For Galileo the 'plane' along which movement would continue indefinitely is a spherical surface. In Newtonian physics this motion is not inertial since it implies the continuous action of a force on the moving body. The ontological status of motion had to be changed from process to state, and the rise of modern science is very much the history of how Aristotelianism had to be overcome in order to achieve this result.

The summary of the dialogue between science and philosophy is also a gross oversimplification. It is too expository and tends to lump together philosophers who have little in common except the fact that they lived at the same time. More space is devoted to the Medieval Schoolmen than to modern philosophers, and although everyone is mentioned, nothing is really said about their contribution to the methods of philosophising as they are practiced today. If there is a lesson we can learn from modern philosophy it is that the moment we put colour and sound and the other secondary qualities in the mind, the primary qualities, known only through the secondary ones, must also become subjective, and the 'whole scientific edifice must collapse' (p. 223). Yet the author fails to inform us why the edifice still looks in quite a good state of preservation.

The Aristotelian paradigm, it is claimed, could provide philosophy of science with the integrating principles it is searching for. What modern scientists say falls straight into the mental slots of Aristotle's philosophy of nature. Aristotle's trivial observation that knowledge begins with vague and undifferentiated notions about the data of experience and that this knowledge is made clearer as we make progress is supported by the 'testimony of psychology' (Piaget), and 'witnesses from science' (Albert Einstein, Ludwig von Bertalanffy, J. Robert Oppenheimer). The hypothetico-deductive method practiced by Aristotle is also used by modern science, and we find 'in the case of both Aristotle and the modern scientists, the reliance on induction with a sufficient number of cases to "see" or "recognise" a general truth' (p. 215). Pace Popper!

More startling still is the claim that the Aristotelian can give a physical reason for the indeterminism in quantum physics. This turns out to be primary matter, identified with what Whitehead terms 'creativity' (p. 242), and indeterminate enough to be correlated with the relativistic notion of field. But what is meant by the indeterminism of quantum mechanics? There have been conflicting views on this subject. Some retain the distinction between the real and the apparent, with real now being the microscopic and random, while the merely apparent is the macroscopic in which classical laws seem to be verified. Others reject the old imaginable particles and waves and favour some type of conceptual symbolism. A third group prefers to argue from the haziness of data and the limitations of measurement to the ultimate unverifiability of classical laws. Again some might want to say that indeterminism is true as a negation of the old determinism, but that it cannot escape the necessity of methodological assumptions and precepts. It cannot succeed therefore in delaying the day when, from a new viewpoint, scientific anticipations once more will envisage some determinate object to be known. What would happen then to primary matter?

The reintroduction of primary matter is accompanied by the reinstatement of the principle of finality 'inductively established by Aristotle in his *Physics*' (p. 70), and vindicated by Whitehead. It is the driving force in biology, and it was used by Galileo in arguing that the heliocentric theory should be preferred because it is simpler (!).

By this time the reader will have made up his mind whether the skeleton still rattles.

What is unfortunate is not Mr Smith's attempt to give an old paradigm a run for its money, for this is quite legitimate. It is the anachronistic interpretation that Aristotle is made to wear, as though he had been philosophising within the conceptual framework of modern science. The ancient and honourable tradition to which the author subscribes is still considered by many as a mummy preserving for all times the science of the Greeks and the common sense of the Middle Ages. Mr Smith does not quite succeed in dispelling this impression.

WILLIAM J. SHEA
Darwin College,
Cambridge

Logic, Methodology and Philosophy of Science. Proceedings of the 1964 International Conference. Edited by Yehoshua Bar-Hillel.
North-Holland Publishing Company, Amsterdam. Pp. viii+440. £4 6s.

It is quite clear from this volume that the Logic, Methodology and Philosophy of Science division of the International Union has been successfully captured by the mathematicians. Of the thirty three papers printed in the Proceedings only two (Hesse, Kyburg) seem to be concerned with the methodology of science, and only one (Mehlberg) with the philosophy of science. I shall review only those that seem to be even remotely connected with the methodology and philosophy of science. Mathematical journals will no doubt deal with the greater part of the Proceedings.

Miss Hesse contributes an interesting attack on the deductivist view of science. Her argument, in short, shows how some actual functions of scientific theories are performed by metaphorical redescription of the field of the phenomena, and not by theory viewed as covering law. It is a great pity that another very interesting paper along the same lines by Wilfred Sellers, mentioned in the programme, is not printed in the volume.

The first intimations of doubt concerning the propriety of modelling confirmation theory on the classical probability calculus outside Popperian circles, appear in Kyburg's paper which seems to be the only other genuinely methodological paper. The other papers in the section really concern technical points in the development of the mathematics of probability.

The 'philosophy' of science is dealt with by Costa de Beauregard, P. G. Bergmann, Havas and Mehlberg. The first three provide summaries of recent and not so recent work in cosmology, de Beauregard being comprehensive and very interesting, Bergmann's and Havas's both being very sketchy. Only Henryk Mehlberg gives us something to think about with an axiom system for relativistic space-time using only one undefined relation. His paper repays careful study.

The papers in the section called Methodology and Philosophy of Life Sciences are on other subjects, e.g. the logic of quotation marks, the history of the theory of automata, and a programme for the empirical study of 'concept formation'. All worthy topics, no doubt, but . . . !

The volume is beautifully produced, and the Editor is to be congratulated for sparing us all the contributed papers. Let us hope that the decision to print only the invited papers, puzzling though their connection with the official programme is in the case of this congress, spreads to other congresses as well.

R. Harré
Linacre College,
London

REVIEWS

Introduction to Logic and to the Methodology of the Deductive Sciences. By Alfred Tarski.
A Galaxy Book, New York, Oxford University Press, 1965. Pp. 252. $1.85.

This paperback edition of Tarski's famous text will make it even more popular with students. The only change over past editions is that new exercises have been added and the annotated list of suggested readings has been expanded and brought up to date.

For those unfamiliar with the book: There are two parts. The first introduces the axiomatic method in a general way, leading up to the notion of a formal theory and the related ideas of consistency, completeness, independence, and the decision problem. The exposition is informal and does not take up, for example, an actual proof of consistency or completeness for any theory. But at the end of each of the six chapters in this part, there are numerous exercises which serve to bring in many technical details not covered in the exposition. In this way a student who does the exercises will learn to do proofs in the 'sentential' calculus, theory of identity, theory of classes, and theory of relations. The second part builds up two axiom systems for the real numbers and explains how the first is simpler while the second is more convenient to use. In the course of these four chapters, the exercises again serve to introduce the student to simple proofs involving concepts of the system. For an introduction to the deductive method, there is no book which combines accessibility and content to a higher degree than this one.

James Cargile

ANNOUNCEMENTS

The University of Western Ontario announce a symposium entitled "The Methodological Heritage of Newton", to be held on 31 March and 1 April 1967. Speakers include Professors P. K. Feyerābend, F. E. L. Priestley, N. R. Hanson and G. Buchdahl. Inquiries to Mrs O. M. Hitchins, Administrative Assistant, Department of Philosophy, Talbot College, University of Western Ontario, London, Ontario, Canada.

The 14th International Congress for Philosophy will be held in Vienna from 2nd – 9th September 1968. Inquiries to Sekretariat des XIV. Internationalen Kongresses für Philosophie, Universität Wien, Universitätsstrasse 7, 1010 Wien.

CALCUTTA
CENTRAL LIBRARY

ABSTRACTS

Dialectica, 1964, 18, Nos. 1-4

Les Antinomies en Droit

Ch. Perelman, ' Avant-propos '

W. Malgaud, ' Les antinomies en droit; à propos de l'étude de G. Gavazzi '

P. Foriers, ' Les antinomies en droit '

L. Morgenthal, ' Les antinomies en droit social '

L. Silance, ' Quelques exemples d'antinomies et essai de classement '

R. Vander Elst, ' Antinomies en droit international privé '

J. Hoeffler, ' Les antinomies en droit public '

G. Boland, ' Quelques propos sur les antinomies et pseudo-antinomies, en particulier en droit administratif '

Ch. Huberlant, ' Antinomies et recours aux principes généraux '

N. Bobbio, ' Des critères pour résoudre les antinomies '

J. Miedzianogora, ' L'adage " Nemo auditur propriam turpitudinem allegans " et le Code Napoléon '

J. Salmon, ' Les antinomies en droit international public '
P. Foriers, ' Les antinomies entre dispositions de droit communautaire et dispositions de droit interne '

I. Tammelo, ' Tensions and Tenebrae in Treaty Interpretation : " Recônnaissance of a Battlefield for Antinomies " '

I. Szabó, ' Des contradictions entre le droit des différents systèmes sociaux '

H. Buch, ' Conceptions dialectiques des antinomies juridiques '

Ch. Perelman, ' Les antinomies en droit. Essai de synthèse '

ABSTRACTS

Dialectica, 1965, 19, Nos. 1/2

La modernisation de l'enseignement mathématique

A. Wittenberg et F. Gonseth, ' Introduction '

A. Wittenberg, ' Questionnaire '

W. Sörensen, ' Un modèle de réponse '

W. Sörensen, ' A propos de la réforme de l'enseignement des mathématiques '

F. Gonseth, ' L'Homo Phenomenologicus '

Il faut bien distinguer dans cet article entre le but et les moyens. Les moyens sont liés à un exemple particulier, celui de la couleur en tant que couleur pour un sujet. Il comporte une analyse méthodologique des procédures utilisées pour parvenir à la connaissance du corps des couleurs virtuelles dont un sujet est porteur dans sa subjectivité, une étude sommaire des résultats auxquels ces procédures conduisent et enfin un dépassement de l'exemple particulier par comparaison, analogie et généralisation.

L'analyse méthodologique conduit à la conception d'une méthode mixte dans laquelle la recherche orientée vers le monde phénoménal (le monde dit physique) et la recherche orientée vers l'univers phénoménologique (celui des formes et des structures de la subjectivité) ont à se prêter un mutuel appui. L'auteur insiste particulièrement sur les procédures tournées vers le sujet, procédures qui rendent possible une connaissance progressive du sujet par lui-même. La recherche procède à la fois par la voie de l'intimité et par celle de l'altérité, dont chacune ne prend tout son sens que par son rapport avec l'autre.

L'étude sommaire des résultats écarte l'idée d'un corps des couleurs identique pour tous les sujets à la fois. Elle suggère au contraire celle d'un corps propre à chaque sujet. Ce corps est à envisager comme une structure de l'univers de la subjectivité. Le conception d'une structure universelle doit faire place à celle de structures individuelles, celles-ci se groupant autour d'un cas normal. L'existence de graves anomalies est soulignée et l'auteur en interprète les conséquences.

Le dépassement du cas particulier se fait en remarquant que la méthode invoquée peut être généralisée, que certaines analogies s'imposent, tandis que d'autres réclament un effort d'imagination. Le but est de parvenir à la conception de l'ensemble ordonné et coordonné de toutes les structures de la subjectivité, ensemble pour lequel le nom d'*Homo phenomenologicus* est proposé. L'auteur dit un mot de la façon dont il conçoit le rapport de ce dernier à l'homme somatique d'une part, à la personne d'autre part.

Sur la voie ainsi tracée, la recherche phénoménologique et la recherche phénoménale sont complémentaires. Leur lien ne peut pas être tranché sans les compromettre l'une et l'autre. Ainsi comprise, la recherche phénoménologique se trouve ramenée au niveau de la recherche scientifique avec laquelle elle forme alors un tout organique.

H. König, ' Über die Grenzen der exakten Erfassung der Qualität " hell " '

ABSTRACTS

A. Bonnet de Viller, ' Science et philosophie '

In former times there was no partition between science and philosophy. What about today? Will philosophy merge in the growing science or will it be able to claim its own living space ? If such could be the case, how are their corresponding fields to be clearly outlined? Is it to be assumed that on both sides of a common border there will always be two separate fields of influence? If not, how is the existing relation between them to be represented?

The answer is provided by the analysis of objective study and its reaction. On this basis a philosophy can be outlined for which scientific knowledge is no longer considered an antagonist, but a partner. Prejudice can and must be eradicated by self-criticism. In former times, save for philosophy, science would not have existed. Without science philosophy in turn can no longer occupy the position it deserves and must be given.

H. Guggenheimer, ' The psychology of real numbers '

The paper contains a discussion of the relation of the Archimedean axiom and completeness from Euclid to the 17th century, and some remarks concerning the discovery of transcendental numbers. Implicitly it contains the thesis that any " horizon de réalité " in Gonseth's sense can be recognised as such only if it has been surpassed.

G. Nádor, ' Scientia mirabilis—Descartes et Leibniz '

J. F. Staal, ' E. W. Beth '

Dialectica, 1965, 19, Nos. 3/4

L'épistémologie de la physique quantique

M. Bunge, ' Physics and reality '

A semantical and methodological analysis of physical theories is performed in order to find out their relation to reality and to human experience. It is shown that every physical theory refers immediately to an idealised model of what is supposed to be a piece of reality—the mediate referent of the theory. Two kinds of physical interpretation of physical symbols are distinguished : objective and operational, and the difference between reference and evidence is stressed. It is claimed that for a theory to be physically meaningful it is necessary that it includes rules of objective reference. In some cases a theory contains, in addition, evidential reference rules, i.e. correspondences between some of its concepts and empirical items. But the test of any theory requires, rather than operational rules, further theories that can bridge the gap between what the theory refers to (e.g. the passage of a cosmic ray) and its remote symptom (e.g. the click of a counter). The conditions for the semantical consistency are sketched and it is argued that there does not exist a semantically consistent interpretation of any of the formalisms of quantum theory. From this analysis arguments in favour of critical realism are drawn.

V. Fock, ' La physique quantique et les idéalisations classiques '

ABSTRACTS

D. Park, ‘The idea of a particle in microphysics’

The development of modern quantum theory has superseded the classical theory of particles. Nevertheless, the imagery and terminology of the classical theory still survive, both in intuitive understanding and mathematical formulation. This paper shows that although the classical concept of a particle may be useful for some purposes of intuition, it is useless for others, and it is never necessary. In microphysics, particles are usually assigned a charge e and a mass m. We discuss the meaning of these quantities in terms of laboratory experience.

G. Emch et J. M. Jauch, ‘Structures logiques et mathématiques en physique quantique’

In this article we present the essential features of modern quantum mechanics in a formalism which resembles in many respects the propositional calculus of classical logic. The purpose is to render explicit the empirical origin of the structure of this calculus, to compare it with the corresponding calculus for classical systems, to sharpen the basic notions of compatibility and complementarity and to sketch some of the more recent insights concerning the properties of states as well as the non-existence of hidden variables.

O. Costa de Beauregard, ‘Le “ paradoxe ” des corrélations d'Einstein et de Schrödinger et l'épaisseur temporelle de la transition quantique’

It is argued that the so-called correlation paradoxes of Einstein-Podolsky-Rosen[1] and of Schrödinger[2] imply that individual quantum processes are connected in time in a way that is symmetric *with* retarded and advanced actions; a ‘fatalistic’ character of the course of events is thus advocated; similar to the one occurring in the so-called ‘Heisenberg picture’ in hyperquantised field theory.

A. Cochran, ‘Life and the wave properties of matter’

G. Gonseth, ‘Sur la méthodologie du calcul des probabilités’

F. Bonsack, ‘En quel sens les probabilités sont-elles subjectives?’

J. L. Destouches, ‘General mathematical physics and schemas, application to the theory of particles’

A. Landé, ‘Non-quantal Foundations of Quantum Mechanics’

La théorie dualistique ne s'impose que par le fait d'une apparente impossibilité d'expliquer la réfraction des électrons autrement que par un intermédiaire ondulatoire et par le fait aussi que le dilemme qu'elle contenait semblait pouvoir être rendu admissible par les artifices de la formulation discursive du positivisme dialectique.

Le première partie de cet article montre que cette façon de voir s'explique par l'omission de certaines lois physiques importantes dont la prise en considération permet au contraire de justifier la réfraction des électrons par une mécanique des seules particules rendant ainsi l' “ idéologie dualistique ” superflue. La seconde partie apporte une explication physique de l'apparence ondulatoire du formalisme quantique. Elle découle de certains principes non quantiques déjà contenus dans la mécanique déterministe, mais que l'on fait servir à l'expression des rapports aléatoires entre les données du niveau atomique.

CORRIGENDUM

In my paper 'A Logic for Evidential Support' (this *Journal*, 17 (1966), 21 ff.) the argument on page 25 is illustrated by wrong and misleading examples.[1] The argument proceeds by considering statements logically equivalent to a universal hypothesis U^1 that are formed by generalising analogously not about the individual elements of U^1's domain of discourse, but about pairs, trios, or *n*-membered classes of these elements, where the domain of U^1 has at least *n* elements. But the generalisations must be understood distributively : not collectively, as the examples in the original version suggested. Such an equivalent for 'All ravens are black' will not be 'All pairs of ravens are pairs of black things', but 'All pairs are pairs of things each of which is a raven only if black'. Similarly, instead of 'If the pair *a*, *b* is a pair of ravens it is a pair of black things' on page 26, I should have written 'The pair *a*, *b* is a pair of things each of which is a raven only if black'. If these alterations are made, the argument is unobjectionable, and the qualification about truth-functional conditionals, on page 25, is no longer required. (A corresponding correction should be made in the argument of 'What has Confirmation to do with Probabilities ?', *Mind*, 75 (1966), 474.)

L. Jonathan Cohen

The Queen's College, Oxford,

---

[1] I am indebted to a conversation with Professor H. Fain for drawing my attention to this.

Revue Internationale de Philosophie

Editorial Board: Profs. Ph. DEVAUX, H. GOUHIER, Ch. PERELMAN and Mrs. Simone SERVAIS

Secretary: 99 avenue de l'Université, Bruxelles 5

Each number is devoted to a particular movement, a particular philosopher, or a particular problem. As complete a bibliography as possible is given.

Truly international in spirit, the *Review* chooses its collaborators among scholars of all countries. The articles are written in English, French, German, Italian, or Spanish.

Publishes 4 numbers annually. (Europe: 350 Belgian francs; other countries: 400 Belgian francs.)

Coming numbers for 1966 and 1967

Maine de Biran (1766-1824)
Leibniz (1646-1716)
Josiah Royce, Franz Brentano, Forme et Pensée

JOURNAL OF THE HISTORY OF IDEAS

A Quarterly devoted to Intellectual History

VOLUME XXVIII JANUARY-MARCH 1967 NUMBER 1

TABLE OF CONTENTS

ARTICLES

Three Ancient Meanings of Matter	*Harold J. Johnson*
From Casuistry to Fiction	*G. A. Starr*
Proudhon and Rousseau	*Aaron Noland*
Modern Views of the Idea of Progress	*W. Warren Wagar*
Problems of Marxist Revisionism	*Charles F. Elliott*

NOTES

Descendit Ad Inferos—Again	*Constance I. Smith*
Thucydides in France: Commynes	*Paul J. Archambault*
The Garden: Marvell's Cartesian Ecstasy	*Daniel Stempel*
Emerson and Indian Philosophy	*Dale Riepe*
Emerson and Empedocles	*Sacvan Bercovitch*

REVIEW

Review of *Aspects of the Eighteenth Century*	*Alfred J. Bingham*

Subscription Rates (outside U.S.A. post paid): $6.50 for one year;
$1.75 for single issue

Payable to

JOURNAL OF THE HISTORY OF IDEAS, INC.
The City College, New York, N.Y. 10031

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences

Edited by ARNE NAESS

Volume 10 *Number* 1 *Spring* 1967

Inquiry as a Transcendental Activity	A. C. Genova
Empirical Certainty and the Theory of Important Criteria	Michael Anthony Slote
On Describing Colours	Bernard Harrison
Can Heidegger be Understood?	Richard Schmitt
Seinsverlassenheit in the Later Philosophy of Heidegger	S. L. Bartky
Discussions:	
Kaufman on Alienation	Paul Diesing and Paul Piccone
Reply to Diesing and Piccone	Arnold S. Kaufman
On Our Knowledge of the Social World	Stuart Silvers
Rejoinder to Stuart Silvers	Arnold Levison
Review discussion:	
R. Bambrough (Ed.): *New Essays on Plato and Aristotle*	Jaakko Hintikka

Published quarterly by Universitetsforlaget, P.O. Box 307,
Blindern, Oslo 3, Norway

Annual Subscription: 35*s*., $6.00, postage included

THE
REVIEW OF METAPHYSICS
A PHILOSOPHICAL QUARTERLY

Edited by RICHARD J. BERNSTEIN

Volume XX, No. 3 *Issue* 79 *March* 1967

ARTICLES. LEGACIES OF THE NINETEENTH CENTURY	
Martin Heidegger	*Who is Nietzsche's Zarathrusta?*
Henry E. Allison	*Christianity and Nonsense*
Arthur Lessing	*Marxist Existentialism*
CRITICAL INVESTIGATIONS:	
Aryeh Kosman	*Aristotle's First Predicament*
Michael Stocker	*Acts, Perfect Duties, and Imperfect Duties*
DISCUSSION:	
John Knox, Jr.	*Blanshard on Causation and Necessity*
BOOKS RECEIVED:	
Susan A. Spooner and Staff	*Summaries and Comments*
PHILOSOPHICAL ABSTRACTS	
ANNOUNCEMENTS	
PROFESSORS EMERITI OF PHILOSOPHY	

Annual Subscription $6.00 Student Subscription $4.00

LYMAN-BEECHER HALL, HAVERFORD COLLEGE, HAVERFORD, Pa. 19041 USA

Readers resident in Great Britain may subscribe at a special rate through our British representative, Professor Stephen Körner, Department of Philosophy, The University, Bristol 8, England

SYNTHESE

An International Journal for Epistemology, Methodology and Philosophy of Science

Editor: JAAKKO HINTIKKA

University of Helsinki and Stanford University

Board of Consulting Editors: David M. Armstrong, Yehoshua Bar-Hillel, Arthur W. Burks Donald Davidson, Hans Freudenthal, Béla Juhos, Jerrold J. Katz, Imre Lakatos, G Nuchelmans, Wesley C. Salmon, Erik Stenius, Patrick Suppes, Klemens Szaniawski H. Tornebohm.

Contents of Volume 16, Nos. 3-4, December 1966

On preference, Probability, and Learning

Exposition of the Theory of Choice under Uncertainty *by Kenneth J. Arrow.*
Strategy and the Logic of Decision *by Joseph D. Sneed.*
Some Formal Models of Grading Principles *by Patrick Suppes.*
Preference Logic and Theory Choice *by Soren Halldén.*
Intrinsic Preferability and the Problem of Supererogation *by Roderick M. Chisholm and Ernest Sosa*
Notes on Preference, Utility, and Cost *by Nicholas Rescher.*
A New Approach to the Formulation and Testing of Learning Models *by Joseph F. Hanna.*
Two Introductions to the Theory of Knowledge (review article) *by S. G. O'Hair*

Reviews

Anatol Rapoport and Albert M. Chammah, Prisoner's Dilemma *by W. G. Runciman.*
Max Black (ed.), Philosophy in America *by Gabriel Nuchelmans.*
Books Received.

Annual Subscription Price Dfl. 48.— post-free

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND

THEORIA

A Swedish Journal of Philosophy

VOLUME XXXII 1966 PART 3

ARTICLES

Per Lindström: On characterizability in $L_{\omega_1\omega}$
Per Lindström: On relations between structures
Per Lindström: First order predicate logic with generalized quantifiers
Gerald C. MacCallum, Jr.: On applying rules
Ernest Sosa: On practical inference and the logic of imperatives
Ernest Sosa: The logic of imperatives

DISCUSSIONS

Roy Harris: Criteria in descriptive semantics
William Todd: Meaning and criteria
Haywood R. Shuford, Jr.: Logical behaviourism and intentionality

MIND

A Quarterly Review of Psychology and Philosophy

edited by Professor **GILBERT RYLE**

CONTENTS FOR ISSUE OF JANUARY 1967

I. On Rational Philosophy of Language: The Programme in Plato's *Cratylus* Reconsidered KUNO LORENZ and JURGEN MITTELSTRASSE
II. Ordinary Language, Common Sense, and the Time-lag Argument R. G. HENSON
III. Action and Causal Explanation GEOFFREY MADELL
IV. Aristotle's Logic of Statements about Contingency A. P. BROGAN
V. Utilitarianism and New Generations JAN NARVESON
VI. Which Word wears the Trousers? S. COVAL and T. FORREST
VII. Memory and Imagination J. O. URMSON
VIII. Discussion Notes:
  More about Motives R. S. PETERS
  Recipes and Causes R. J. HAACK
  The Mysteries of Udolpho: A Reply to Professor Jeffrey and Bar-Hillel K. R. POPPER
  Austin's Worries about 'I state that . . .' C. OLSEN
  The Surprise Exam: Prediction on last day uncertain J. A. WRIGHT
  Mr. Strawson on the Traditional Logic T J. SMILEY
  Professor Bradley's Avowals M. J. CRESSWELL
  "Deontic and Atheoretical Logic" E. SCHUH
  Plato on Moral Principles RAPHAEL DEMOS
IX. Critical Notice:
  The Explanation of Behaviour by Charles Taylor D. W. HAMLYN
X. New Books
XI. Notes
XII. List of Officers and Members of the Mind Association

Annual Subscription 25s, post free; single copy 7s 6d, plus postage

Published for MIND ASSOCIATION by

BASIL BLACKWELL 49 BROAD STREET OXFORD

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

DECEMBER 1966, VOL. XXVII, No. 2

CONTENTS

I. ARTICLES
  Plato's Doctrine of Truth: Orthótes or Alétheia?, by HENRY G. WOLZ
  The Construction of Empirical Concepts, by DONALD S. LEE
  Existence and Time, by K. W. RANKIN
  The Young Marx—A Reappraisal, by DONALD CLARK HODGES
  On Alienation and the Social Order, by IRVING LOUIS HOROWITZ
  Men, Machines, Materialism, and Morality, by PETER T. MANICAS
  On Characterizing the Psychological, by HERBERT HEIDELBERGER
  The Sensitive Term "Pain," by C. R. OCHS
  Austin On Incorrigibility, by JOHN E. ATWELL
  Rules and Exceptions in Ethics and Aesthetics, by ELMER H. DUNCAN
  On the Concept of Universals, by HANS ASRIEL WEISSMANN
  Simple Ideas, by DAVID L. PERRY

II. DISCUSSION
  R. B. Perry on Value and Existence, by BERTRAM JESSUP

III. REVIEWS, NOTES AND NEWS, AND RECENT PUBLICATIONS

Subscription prices $6.00 per year for individuals; $7.50 per year for institutions. Separate issues, $1.65 and $2.00 respectively. Postage added for back issues only. A circular listing the contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo, Buffalo, New York, 14214

CALCUTTA UNIVERSITY
CENTRAL LIBRARY.

Printed in Great Britain at The University Press, Aberdeen